COLOR, SHAPE, AND NUMBER IDENTITY-NONIDENTITY RESPONDING AND CONCEPT FORMATION IN ORANGUTANS

A Dissertation
Presented to
The Academic Faculty

by

Ursula Simone Anderson

In Partial Fulfillment
of the Requirements for the Degree
PhD in the
School of Psychology

Georgia Institute of Technology

December 2011

Copyright © 2011 Ursula S Anderson

COLOR, SHAPE, AND NUMBER IDENTITY-NONIDENTITY RESPONDING AND CONCEPT FORMATION IN ORANGUTANS

Approved by:

Dr. Terry L. Maple, Advisor
School of Psychology
Georgia Institute of Technology

Dr. Marcus Jack Marr
School of Psychology
Georgia Institute of Technology

Dr. Tara S. Stoinski
Zoo Atlanta and
The Dian Fossey Gorilla Fund International

Dr. Greg Corso
School of Psychology
Georgia Institute of Technology

Dr. Nancy Nercessian
School of Interactive Computing
Georgia Institute of Technology

Dr. Michael Beran
Language Research Center
Georgia State University

Date Approved: July 28, 2011

ACKNOWLEDGEMENTS

While completing my dissertation, I was blessed by God to receive support from my friends and family who were there not only during the crying and ranting, but also during the smiling and laughing. I would like to thank the Interlibrary Loan program at Georgia Tech, which provided an invaluable service by electronically delivering 100+ journal articles to me. I also thank Zoo Atlanta for generously providing animal access and resources; in particular, the primate keeper and curator staff for their assistance with the orangutans and the commissary staff for keeping the grapes flowing through 262 days of data collection.

Finally, several organizations deserve recognition for providing funding for my doctoral studies and dissertation. They are the Facilitating Academic Careers in Engineering and Science fellowship program, American Psychological Association's Committee on Animal Research & Ethics and Dissertation Research Award program, Southern Regional Education Board's Doctoral Scholars Program, Hispanic Scholarship Fund, Congressional Black Caucus Foundation, the Churches Homes Foundation, and Beran Charitable and Scholarship Foundation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

The ability to recognize sameness among objects and events is a prerequisite for abstraction and forming concepts about what one has learned; thus, identity and nonidentity learning can be considered the backbone of higher-order human cognitive abilities. This is one reason why it is critical to understand how nonhuman primates learn about identity relations. Furthermore, given the adaptive significance of using concepts, it is important to investigate if and how nonhuman primates form identity concepts for which they categorize or classify the stimuli around them.

Identity concepts involve discriminating among and between objects, physical properties, states, and events in the natural world based on the shared relation of equality (Lock & Colombo, 1996; Spinozzi, 1996; Thompson & Oden, 2000; Tomasello & Call, 1997). One can subdivide the relation of identity into whole- and part-identity (Evans & Smith, 1988; Smith, 1989), with the former characterized by discovering identity relations holistically (e.g., a penny is identical to another penny) and the latter by discovering identity relations between the constituent properties or parts of objects (e.g., a red square is identical in color to a red circle). Color and shape are the two common properties for which objects may be identical in one regard, but not in the other; additionally, number (cardinality) may be considered a constituent property of a collection of objects. Discovering identity relations between the constituent properties of objects is an important ability that often characterizes the comparisons that humans make so it is important to devote attention to understanding how nonhuman primates process and conceptualize part-identity. Because the ability to generalize the results of learning is to what concepts ultimately reduce, the series of experiments herein first investigated

responding to part-identity and -nonidentity and whole-identity and -nonidentity and then explored the generality of such learning to the formation of concepts about color, shape, and cardinal number.

The objectives and aims of Experiments 1, 2, and 3 were threefold: First, establish and evaluate concurrent whole-identity and -nonidentity responding to stimuli that differed only in one regard, their color, using for the first time a novel conditional discrimination procedure that was akin to the S/D discrimination task. Second, evaluate performance differences between concurrent color and shape part-identity responding with a larger set of colors and shapes and then do the same for concurrent color and shape part- and whole-identity responding. Finally, after the conditional discrimination was established, evaluate color and shape part- and whole-identity concept formation by introducing novel problems that featured a small set of novel colors and novel shapes.

The data from Experiments 1, 2, and 3 indicated that the two orangutans learned to respond concurrently to color whole-identity and -nonidentity and they responded faster to color whole-identity. Additionally, both subjects learned to respond concurrently to color and shape part- and whole-identity and for the most part, it was easier for them to do so with color part- and whole-identity problems than shape part- and whole-identity problems. Further, their learned responses to color and shape part- and whole-identity fully transferred to novel color part-identity problems for both subjects and fully transferred to novel color and shape whole-identity problems for one orangutan.

The objectives and aims of Experiments 4, 5, and 6 were also threefold: First, establish concurrent identity and nonidentity responding to a small set of numerosity stimuli that had cue-ambiguous irrelevant dimensions for the first time in orangutans.

Second, establish identity responding to a small set of numerosity stimuli that had cue-ambiguous irrelevant dimensions before establishing the same when one or more irrelevant dimensions were cue-constant. Finally, after the conditional discrimination was established, investigate the extent to which numerical identity concepts about cardinal number formed by introducing novel and familiar numerosities that were instantiated with novel and familiar cue-constant and cue-ambiguous element colors and shapes.

The data from Experiments 4, 5, and 6 showed that one subject learned to judge numerical identity when both irrelevant dimensions were cue-constant, but the subject did not do the same when one or more irrelevant dimensions were cue-ambiguous. Further, the subject's accuracy was affected by the numerical distance and the numerical total of comparisons during acquisition of the conditional discrimination. The subject subsequently formed a domain-specific concept about numerical identity as evinced by the transfer of learning to novel numerosities instantiated with novel, cue-constant element colors and shapes and novel numerosities instantiated with cue-constant, familiar element colors and shapes.

This dissertation provided evidence about the extent to which orangutans learned to respond to color, shape, and number identity and nonidentity and subsequent concept formation from such learning. The findings from this study will help in understanding the convergence and divergence in the expression abstraction in the primate phylogeny, thus, informing our understanding about the origins and mechanisms of cognition in human and nonhuman primates.

CHAPTER 1

INTRODUCTION

Beginning in early childhood, humans discriminate patterns among objects, events, and ideas across the domains of time, space, and the sensory modalities (Bornstein, 1984; Holyoak, Gentner, & Kokinov, 2001; Oden, Thompson, & Premack, 1988; Tomasello, 1999). Discriminating patterns is not the same thing as forming concepts about those patterns, but the two processes share a common activity—assigning entities to categories (Neisser, 1987). In particular, discrimination is a prerequisite for abstraction and forming concepts (Holyoak, Gentner, & Kokinov, 2001) as behavior can only be labeled conceptually-mediated if one learns to discriminate between and among classes and that behavior generalizes to novel instances (Thompson, 1995). Concepts give the world stability and are functionally adaptive and cognitively economic for those who use them. They capture the idea that many things are similar in some respect. Once an entity is assigned to an existing class, then some of its attributes can be inferred and it can be thought about and responded to in ways already mastered. Moreover, by virtue of membership in a familiar class, behavioral adjustment to novel entities with which one has no prior experience or reinforcement history is possible, which reduces the amount of information that needs attention and remembering (Smith, 1981; Thompson, 1995; Thompson & Oden, 1995, 2000; Tomasello & Call, 1997; Zentall, Wasserman, Lazareva, Thompson, & Rattermann, 2008).

William James (1890/1981) noted more than a century ago that “sense of sameness is the very keel and background of our thinking” (p. 434). Today, concepts of identity and nonidentity, which involve discriminating among and between entities based on the shared relation of equality (Lock & Colombo, 1996; Spinozzi, 1996; Thompson & Oden, 2000; Tomasello & Call, 1997), are the most commonly studied type of relational concepts. Given its pervasiveness as a research topic in humans, it is not surprising that identity-nonidentity relational learning and concept formation are considered fundamental to understanding nonhuman primate cognition (Lock & Colombo, 1996; Oden, Thompson, & Premack, 1988; Spinozzi, 1996; Thompson & Oden, 2000; Tomasello & Call, 1997).

What of number? Number—an abstraction that represents how many discrete things are perceived and the only property of a set that remains invariant when other characteristics of the set change (Piaget, 1941/1965)—not only fills our world, but also fills the world of primates. Research in human numerical cognition focuses on how, if, and when humans use number as a salient dimension, with the findings indicating that the ability to perceive numerical information begins in infancy (Cooper, 1984; Gelman & Gallistel, 1986; Strauss & Curtis, 1984). Likewise, more than a century's worth of research has determined that primates process and use numerical information in a variety of ways (for reviews, see Boysen, 1997; Boysen & Capaldi, 1993; Boysen & Hallberg, 2000; Davis & Memmott, 1982; Gallistel & Gelman, 2000; and Tomasello & Call, 1997).

Studying higher-order cognitive abilities like relational learning, concept formation, and numerical cognition in nonhuman primates is an attempt, for one, to relate their abilities to the abilities of humans (Himes, 1999). If one conceptualizes cognition as covert behavior, then according to the basic rules of evolution, cognitive abilities are subject to natural selection like any biological characteristic, and patterns in the expression of behavior and cognitive abilities should exist across species with respect to an organism's phylogeny. Because the taxonomic relationships among humans and nonhuman primates are known (e.g., humans and chimpanzees, bonobos, gorillas, and orangutans are closely related sister species that share a common ancestor that dates back 5 to 10 million years ago), there is an evolutionary basis for applying comparative methods to evaluate patterns of divergence and convergence in cognitive abilities (Haun, Jordan, Vallortigara, & Clayton, 2010; Parker, 1999).

Discovering the areas of cognitive generality and specificity between humans and primates is complicated because each species has its own specific limitations (Wright, 1992), different methodological tools employed across species may lead to task-specific differences (Hauser, 1997), and schooling and language learning may mask or distort shared mechanisms (Fabre-Thorpe, 2001; Geary & Lin, 1998). Further, human-like language abilities are precluded in all but a few animal species (e.g., language-trained apes) so language cannot yet be assigned a significant role in the behavior of nonhuman animals in natural settings (Lattal & Perone, 1998). In light of the aforementioned, the most important comparisons may be between nonhuman primates and nonverbal human infants (Hauser, 1997).

Given the adaptive significance of using concepts, it is important to investigate if and how primates form concepts for which they categorize or classify the stimuli around them (Thompson, 1995; Thompson & Oden, 1995, 2000; Tomasello & Call, 1997). One problem in investigating concept formation is that psychologists often talk about concepts, but then only study categorization. Further, most researchers assume that concept formation is pure and simple, but there are enough differences in the nature of the input (c.f., random dot patterns vs. geometric shapes, multi-item arrays vs. two-item arrays, two vs. three dimensional stimuli, etc.) to warrant close differentiated examinations (Mandler, 2000).

Investigations about the extent to which apes, monkeys, and human infants form conceptual understandings of identity and nonidentity about color, shape, and cardinal number from identity and nonidentity judgments about color, shape, and cardinal number are limited. The aforementioned trend in the literature is in comparison to investigations about relational learning and concept formation that involves objects and continuous quantity (e.g., the combined influence of number and length, volume, area, etc.). A comparative examination among apes, monkeys, and human infants may contribute to our understanding about when higher-order cognitive abilities emerged in the phylogenetic scale, how they have increased in complexity since their emergence, and the extent to which they depend upon language (Fabre-Thorpe, 2001; Hauser, 1997; Pearce, 1988).

Proper application of the comparative method, however, requires knowledge about the abilities of all sister ape species, especially those that are nonverbal given that researchers theorize that language promotes a kind of flexible thinking that leads to the development of higher levels of conceptualization (Fabre-Thorpe, 2001; Hauser & Carey, 1998; Kotovsky & Gentner, 1996; Parker, Mitchell, & Miles, 1999; Pearce, 1988; Premack, 1988; Thompson & Oden, 1995; Tomasello, 1999). The majority of literature concerning concept formation about color, shape, and number in apes, though, is explored in language-trained chimpanzees. This dissertation, thus, attempts to balance the scale by providing evidence about the extent that nonlanguage-trained orangutans learn to judge color, shape, and cardinal number identity-nonidentity and form concepts of identity-nonidentity from such judgments. As such, this line of investigation informs our understanding about the origins and mechanisms of cognition in humans and nonhuman primates.

CHAPTER 2

CONCEPTS

In the human cognition literature, a concept is regarded as a mental representation of a category, which need not have a real world counterpart, or as knowledge that facilitates categorization and a category is defined a class of objects or events that belong together. Cognitive psychology focuses on the structure and organization of concepts and primarily describes terminal performance (Medin, 1989; Smith, 1981). One way to think about the relation between concepts and categories is to describe concepts as a basis or rule for categorization, which is a differential response to a set of stimuli (Herrnstein, 1990; Medin, 1989; Smith, 2005). Thus, concept formation and categorization are intimately connected to the extent that theories about conceptual structure are not separated from those proposed for categorization.

Recasting terminology and taxonomy in terms of the conditions necessary and sufficient for conceptual behavior, the behavior analytic perspective takes a functional approach by examining what types of behavior are conceptual and how conceptual behavior emerges (Astley & Wasserman, 1996; Palmer, 2002; Wasserman & DeVolder, 1993; Zentall, Galizio, & Critchfield, 2002). In this regard, Keller and Schoenfeld's (1950) outline serves as the standard, "when a group of objects gets the same response, when they form a class the members of which are reacted to similarly, we speak of a concept" (p. 154) and "generalization within classes and discrimination between classes—this is the essence of concepts" (p. 155). Further, it is common to speak of a class of stimuli (i.e., a stimulus class) that occasion a common response in a given context as a category or concept.

In nonhuman animals, the study of concept learning is rooted in the field of comparative psychology, with recent work integrating behavior analysis and learning theory (e.g., discrimination, generalization, and laws of learning), cognitive psychology (e.g., memory, attention, and representation), and evolution and ecology (e.g., adaptation and ecological relevance). Animal cognition, the area of inquiry that lies at the intersection of human cognition and animal learning, draws on all of the aforementioned research traditions in its approach to the study of concept formation (Terrace, 1984; Thompson, 1995; Thompson & Oden, 2000; Tomasello & Call, 1997; Zentall & Smeets, 1996). Animal cognition researchers investigate conceptually-mediated behavior, that is, whether nonverbal organisms are capable of discriminating between and within common classes of objects, states, and events and across multiple domains in ways that do not follow the simple associative processes often attributed to animals (Lock & Colombo, 1996; Thompson & Oden, 2000).

It should be clear by now that the terms concept, category, stimulus class, and their derivatives are used interchangeably and loosely across and within research domains. One must also add the term abstraction to the ambiguous terminology and taxonomy list. An abstraction is a discrimination based on a *single* property of stimuli, independent of other properties, thus, it involves generalization among all stimuli with that property (e.g., all red stimuli as opposed to specific red objects). The term concept is sometime used to refer to abstraction involving more than a single property, thus, it involves generalization among stimuli along multiple dimensions (e.g., humans as opposed to orangutans). More often the terms abstraction and concept are used interchangeably (Catania, 1998). In any case, the taxonomy and terminology that fall under the rubric of concept formation are not the important issue; instead, it is important to investigate whether animals learn or form concepts or stimulus classes for which they categorize or classify the stimuli around them (Thompson, 1995; Thompson & Oden, 1995, 2000; Tomasello & Call, 1997).

2.1 Inferring Concept Formation

> Before I am prepared to say that a pigeon has a concept of "person", I want to know whether what I teach it about …"person," say a tall man wearing a white coat, it will generalize to another instance of the same concept, say a short woman wearing a maroon sweater—and that it will generalize in this particular way rather than to all white objects (e.g. refrigerators or laboratory doors), or all objects 179 cm tall (Lea, 1984, p. 271).

To investigate whether animals learn or form concepts or classes for which they categorize or classify the stimuli around them, as Lea (1984) indicated, it is not sufficient merely to demonstrate discriminative responding in the face of stimulus variability and then conclude that an organism's behavior was conceptually-mediated. Concept formation is inferred when the same response is spontaneously occasioned by many stimuli in a set (or by the relations among the stimuli in a set) and that response must occur at a much lower probability (or at a much lower rate) in the presence of stimuli that are not members of the class, in other words, generalization within and discrimination between classes (Fields & Reeve, 2000; Herrnstein, 1990; Keller & Schoenfeld, 1950; Lea, 1984).

2.1.1 Operant Conditioning Procedures

Discrimination learning is a prerequisite for forming concepts (Holyoak, Gentner, & Kokinov, 2001). Inferences of concept formation usually involve assessing differences in responding in the presence of different stimuli, which is a stimulus control issue (Saunders & Williams, 1998). Stimulus control refers to any difference in responding in the presence of different stimuli. Simple and conditional discrimination procedures are instrumental in studying higher-order processes like concept formation because the study of concept formation requires a training history in which the same response is reinforced in the presence of a number of stimuli that contain the element or elements to which control is being established.

In the prototypical simple discrimination task, a reinforcer is delivered after responses to positive discriminative stimuli (S+), but not after responses to negative discriminative stimuli (S-). For example, in the oddity task, subjects are simultaneously presented with sets of three or more stimuli (e.g., AAB or ABA, with the letters of the alphabet signifying different stimuli) and selection of the odd stimulus is rewarded regardless of its spatial position (Saunders & Williams, 1998; Zentall, Hogan, & Edwards, 1984).

In conditional discrimination, the function of a discriminative stimulus, whether S+ or S-, changes based on the presence of another stimulus, the conditional stimulus; in other words, the simple discrimination that is reinforced depends on which conditional stimulus is presented (Saunders & Williams, 1998). The same/different (S/D) task is an example of the conditional position discrimination procedure. In this task, for example, subjects learn to associate one configuration of stimuli (identical stimuli: AA, BB, and CC) with pressing the left button because this response is followed by reinforcement and another configuration of stimuli (nonidentical stimuli: AB, BC, and CD) with pressing the right button because this response is followed by reinforcement (Nakagawa, 2003; Saunders & Williams, 1998; Zentall, Hogan, & Edwards, 1984). Matching-to-sample procedures are another prototypical conditional discrimination procedure. In the two-choice matching-to sample (MTS) task, a subject is presented with a single sample stimulus and two comparison stimuli, and the subject receives differential reinforcement for choosing the comparison stimulus that matches the sample stimulus (Saunders & Williams, 1998; Zentall, Hogan, & Edwards, 1984).

What about the simple and conditional discrimination procedures employed with human infants? The conjugate reinforcement paradigm requires that infants activate reinforcement (auditory, visual, nonnutritive, or social) with their arm or leg movements, with the intensity, rate, or salience of reinforcement contingent on the infant's behavior. For example, with conjugate mobile reinforcement, infants lie supine in cribs with a cord attached from their leg to a mobile that hangs overhead such that their movements (the operant) activate the mobile. If infants who have learned to associate their kicks with activating the mobile during a training phase continue to kick at a high rate when tested with the same mobile under nonreinforcement and their rate of kicking does not exceed baseline rates when tested with a novel mobile, then it can be said that they discriminate the details of the novel mobile from the details of the training mobile (Colombo, 1993; Piek, 2006; Rovee-Collier, 1996; Weisberg & Rovee-Collier, 1998).

Finally, the visual anticipation paradigm is a method that has both operant and classical conditioning features. With this procedure, infants learn spatial, spatiotemporal, or other kinds of patterns among stimuli (a predictable stimulus order or movement) such that they start to anticipate the pattern. For example, infants are seated in front of video displays while a target stimulus appears in the center of the screen, then disappears, and then is followed by visual reinforcement that appears on the right side of the screen if a red target stimulus was presented or on the left side of the screen if a blue target stimulus was presented. An infant's typical response is to make a head turn or visually fixate on stimuli that appear on the screen in the different locations, but the question is whether infants are capable of predicting or anticipating patterns, especially when novel members of the learned class are introduced during transfer testing.

Of final note, the stimuli employed in simple and conditional tasks may vary in the number of dimensions they share with each other. Relevant dimensions distinguish odd from nonodd stimuli or matching from nonmatching stimuli and irrelevant dimensions are those that do not. When operant procedures are used, the relevant dimension is the property that is correlated with differential reinforcement. Irrelevant dimensions may be of two types: those that are cue-constant are shared by all stimuli and those that are cue-ambiguous differ among all stimuli in an uninformative way. Thus, irrelevant dimension cue constancy and ambiguity are measures of between-stimulus heterogeneity or variability. In a three-choice oddity task, for example, a blue circle, red circle, and red circle would respectively serve as the odd and nonodd stimuli if color was the relevant and shape the irrelevant cue-constant dimension. In a three-choice S/D task, a blue triangle and red triangle versus a green square and yellow circle would respectively serve as the identical and nonidentical pair if shape was the relevant dimension and color the irrelevant cue-ambiguous dimension. When problems are constructed such that nonodd (or matching) stimuli are identical, like in the first example, the task is called a simple or conventional task. When problems are constructed such that the matching (or nonodd) stimuli are not identical, but share more properties with each other than they share with the odd stimulus, like in the latter example, the task is called a dimension-abstracted task (Bernstein, 1961; Noble & Thomas, 1985; Steirn & Thomas, 1990; Strong, Drash, & Hedges, 1968; Thomas & Frost, 1983).

2.1.2 Transfer Tests

Mastery of operant tasks does not presume the formation of concepts. Assessing transfer is the standard method used to determine if behavior goes beyond simple associative and perceptual processes after operant procedures have established discriminative responding. Transfer of learning tests measure the degree to which organisms recognize class memberships when comparing stimuli by demonstrating that the effects of contingencies applied to some stimuli of a given class (i.e., baseline performance to training stimuli) generalize or shift to other members of the same class (Saunders & Williams, 1998; Spinozzi, 1996; Thompson, 1995; Thompson & Oden, 1995, 2000; Tomasello & Call, 1997; Zentall, Galizio, & Critchfield, 2002). In addition to transfer of learning tests, the transfer may in the form of generalization of function or in the emergence of untrained relations from relations that are explicitly trained (see Zentall, Galizio, & Critchfield, 2002).

To demonstrate conceptual behavior, transfer of learning tests must use novel exemplars to ensure that there are no features other than the concept facilitating transfer. If novel stimuli are not used then transfer performance could be controlled by extraneous factors, the features of training stimuli, reinforcement history, or background cues that parallel the concept (Herrnstein, 1990; Jitsumori, Siemann, Lehr, & Delius, 2002;

Spinozzi, 1996; Thompson, 1995; Thompson & Oden, 2000; Wright & Katz, 2006; Zentall, Galizio, & Critchfield, 2002). Novel exemplars may be novel in the sense that the stimuli have never been seen before by subjects or task-novel in that they were not used during acquisition or transfer testing for the task-at-hand or for a similar previous task. There are also semi-novel exemplars, sets that contain at least one stimulus that was used in training or transfer testing before being employed in the task-at-hand. For example, a familiar-novel pair of stimuli contains one stimulus that was used in training and one novel stimulus. In many cases, transfer tests with semi-novel exemplars cannot show concept learning, but sometimes they can and do.

Further, with simple and conditional discrimination procedures, it is critical that transfer tests not involve differential reinforcement of correct responses to novel exemplars because it is then possible that performance reflects acquisition of a learning set in which an initially narrowly construed association was applied to an increasingly broader class of objects. If one chooses to reinforce responses differentially during transfer testing then one-trial (also called trial-unique) presentations of novel problems is preferable to prevent subjects from rapidly learning how to respond accurately. If the same novel exemplars within different problems or the same problem is repeatedly presented under differential reinforcement, then one should evaluate transfer of learning using first- or early-trial performance for each problem or for the transfer test session.

Although transfer of learning assessments remains the critical test for concept formation, there is no consensus about what constitutes successful transfer of learning. Some psychologists argue that learning must be relatively faster during transfer when presented with novel exemplars than it was during baseline to infer conceptual behavior.

If learning is at the same basic rate during training and transfer then the inference is that a concept has not been learned (Tomasello & Call, 1997).

Some psychologists propose that the formation of a concept is confirmed by the maintenance of a high level of accuracy during the transfer test phase when trained animals are presented with novel exemplars. This may mean that performance with novel stimuli during transfer testing is equivalent to baseline performance, is better than what chance would predict, or is better than chance but below baseline performance (Spinozzi, 1996; Vonk & MacDonald, 2002; Wright & Katz, 2006; Zentall, 1996; Zentall, Hogan, & Edwards, 1984). Wright and Katz (2006) termed partial transfer as when transfer performance is better than what chance would predict, but below baseline performance and full transfer as when transfer performance is equivalent to baseline performance and above what chance would predict. Partial transfer is not as encouraging as is full transfer because it is possible that multiple cues other than the concept are controlling behavior. In any case, the immediate successful transfer of discriminative performance to novel exemplars (i.e., early-trial learning) instead of successful transfer of discriminative performance to novel exemplars across repeated presentations, is the most compelling evidence for concluding conceptual behavior because it suggests that learning did not function in a simple associative manner during acquisition (Thompson & Oden, 2000).

2.2 Types of Concepts

Three broad attributes—perceptual, associative, and relational—seem to unite events within a category or stimuli class (Zentall, Galizio, & Critchfield, 2002). Although this dissertation will use the aforementioned system to type concepts, other researchers use different systems. For example, concepts may also be grouped as similarity- and nonsimilarity-based, in other words, stimulus classes for which the defining characteristic is similarity versus the converse, nonsimilarity (Wasserman & DeVolder, 1993; Zentall, 1996). Because perceptual and associative concepts are only tangential to this dissertation, I present only a brief discussion of these topics to establish a framework for thinking about relational concepts.

2.2.1 Perceptual Concepts

Perceptual concept learning is related to the most familiar form of categorization in humans. Perceptual concepts involve sorting stimuli into appropriate categories based on perceptual identity and similarity, in other words, grouping stimuli that share one or more physical properties into classes. Humans would describe members of a perceptual class in terms of a category label like ‘red’ things. The stimuli that compose perceptual classes cannot be defined by a few simple features, but instead possess many features that control categorization so it is often difficult to specify the particular common elements that might be used to classify category members from nonmembers.

Investigations of perceptual concept formation tend to focus on natural concepts. Natural concepts are those that form in relation to stimuli that occur in the natural environment (e.g., people, trees, and animals). The stimuli within natural categories are physically more similar to each other than they are to stimuli from different categories (Herrnstein, 1984; Jitsumori & Delius, 2001; Smith, 1981; Thompson, 1995). Because perceptual similarity often guides responding, this type of concept learning is considered a basic-level conceptual behavior largely under the control of the behavioral principles of stimulus discrimination and generalization (Katz, Wright, & Bodily, 2007; Zentall, Galizio, & Critchfield, 2002; Zentall et al., 2008).

2.2.2 Associative Concepts

Associative concepts involve forming categories that are made up of arbitrary stimuli deemed to be equivalent because they are associated with a common event, response, or outcome as the stimuli within classes bearing no obvious physical similarity to each other (Katz, Wright, & Bodily, 2007; Zentall et al., 2008). Thus, the formation of associative concepts attests to the ability to produce novel and appropriate responses in the presence of physically dissimilar objects and events.

Equivalence classes are a specific type of associative concept. Sidman (1990, 1997) defined stimulus equivalence as an asset of stimuli that all bare the same relation to one another if all of the tests for the properties of equivalence—reflexivity, symmetry, and transitivity—are satisfied. One way to think about these three relations is to say that two ideas that were never associated directly with each other, but were each associated with a third idea, could potentially come to be associated such that the three ideas become interchangeable (Catania, 1998; Green & Saunders, 1998; Zentall, 1996).[1]

Functional (equivalence) classes are also a specific type of associative concept. A functional class is composed of a group of discriminative stimuli that all control the same behavior. One can conclude that stimuli are functionally equivalent when a variable applied to one, like a change in contingencies, is sufficient to similarly change behavior in the others (Clayton & Hayes, 1999; Sidman, 1990; Sidman, Wynne, Maguire, & Barnes, 1989; Urcuioli & Zentall, 1993; Zentall, 1996; Zentall, Clement, & Weaver, 2003; Zentall, Galizio, & Critchfield, 2002; Zentall et al., 2008).

---

[1] There is also the relational frame account of stimulus equivalence (for review, see Hayes, Barnes-Holmes, & Roche, 2001).

2.2.3 Relational Concepts

Relational (also called abstract) concepts are illustrated when membership in a common class is based on the relations among stimuli, which transcends the individual features of stimuli. To form a relational concept, an organism must be able to isolate relevant relational invariants from the physical properties of the entities engaged in the relation and transfer that understanding to novel arguments. In other words, relational concept learning involves judging the nonphysical or abstract relationship (e.g., temporal, perceptual, and spatial) between stimuli based on a rule that transcends any particular set of exemplars. Example types of relational classes are ‘larger than’ (i.e., larger-smaller), ‘more than’ (i.e., more-less), and ‘identical to’ (i.e., identity-nonidentity). Thus, a single stimulus can belong to one class (e.g., better than) if the stimulus to which it is compared is worse than it and another class (e.g., worse than) if the stimulus to which it is compared is better than it (Katz, Wright, & Bodily, 2007; Thompson, 1995; Thompson & Oden, 2000; Tomasello & Call, 1997; Wright, 1992; Zentall, Galizio, & Critchfield, 2002; Zentall et al., 2008).

Relational concepts are said to be a prerequisite for, or the basis of, higher-order cognitive skills like understanding mathematical equivalence operations, analogy, and Piagetian conservation (Czerny & Thomas, 1975; Thomas & Peay, 1976). Understanding social relationships, and possibly foraging, seems the likely impetus for the evolution of relational concepts because these skills depend on comparing conspecifics and food sources (Hauser, 1997; Tomasello & Call, 1997). Identity and nonidentity are the most commonly studied relational concepts. Identity concepts involve discriminating among and between objects, physical properties, states, and events in the natural world based on the shared relation of equality (Lock & Colombo, 1996; Spinozzi, 1996; Thompson & Oden, 2000; Tomasello & Call, 1997). The identity relation may be subdivided into whole- and part-identity. Whole-identity is characterized by discovering identity relations holistically across all dimensions and part-identity by discovering identity relations between the constituent properties or parts of objects (Evans & Smith, 1988; Smith, 1989). Discovering identity relations among the constituent parts of objects is an ability that often characterizes the comparisons that humans make, however, less research attention is devoted to investigating processing and conceptualization of part-identity than whole-identity.

Color part- and whole-identity MTS tasks require that color, the mix of light wavelengths leaving the surface of objects, be the relevant dimension that distinguishes whether stimuli are identical or nonidentical. Likewise, shape part- and whole-identity MTS tasks require that shape, the region bounded by continuous contour, be the relevant dimension. Furthermore, for part-identity MTS tasks, at least one irrelevant dimension must be cue-ambiguous (i.e., its instances differ among all stimuli being compared in some uninformative way); for example, a green square is identical to a red square, not a yellow rectangle in terms of shape part-identity. For whole-identity MTS tasks, all irrelevant dimensions are cue-constant (i.e., its instances are shared by all stimuli being compared); for example, a green square is identical to a green square, not a yellow square in terms of color whole-identity.

The next sections detail the empirical literature about color and shape identity and nonidentity responding and concept formation in apes, monkeys, and nonverbal human infants. Nonverbal human infants were defined as infants 12 months of age or younger because infants typically produce their first word around one year of age (Benedict, 1979; Berger, 2008; Lamb, Bornstein, & Teti, 2002; Menyuk, Liebergott, & Schultz, 1995). Second, because of the scope of the dissertation research, only empirical studies using operant or associative learning methods in which identity-nonidentity responding as well as subsequent concept formation could be assessed are discussed. Finally, it was useful to exclude studies that involved two or more relevant dimensions, cross-modal and extradimensional concept formation, and symbolic language systems (i.e., lexicons and lexigrams).

I limited my evaluations of concept formation to early-trial transfer performance, which I set at 25 or fewer trials of a single problem when differential reinforcement was employed.[2] Further, statistical analyses to determine whether baseline and transfer test performance differed significantly from each other or differed from chance were not performed in many studies described herein. For these studies, I considered baseline and transfer test performance equivalent when their accuracy differed by 5% to 6% or less and I considered transfer test performance at what chance would predict if accuracy differed from chance by 15% to 16% or less.[3] Finally, Douglass (1925) recognized that there is no limit to a concept's extension or perfection; instead, there are varying degrees in the attainment of a perfect concept. Based on this idea, it was useful in this dissertation to distinguish the level at which concepts formed according to which dimension's

---

[2] Data that showed that a concept *did not* form across more than 25 differentially reinforced trials of a single problem were considered for inclusion

[3] When statistical evaluations were absent, chance probabilities for correct responses were estimated to be .50 for two-choice simple and conditional discriminations, .33 for three-choice simple and conditional discriminations, and so forth unless otherwise indicated in the text.

instances were made novel for the transfer test stimuli (Roberts, Robbins, & Everitt, 1988). The levels of concept formation that are proposed herein may be analogous to the levels of relational knowledge proposed by Smith (1984) and this possibility is presented in the Discussion section at the end of Chapter 2.

The lowest level of concept formation was assessed by introducing novel instances of the irrelevant dimension; for example, training responding in which shape is the relevant dimension and transfer testing with the familiar training shapes instantiated in novel colors. If an organism maintains discriminative responding to novel instances of the irrelevant dimension then it has demonstrated a conceptual understanding that the identity relation applies to specific trained members within the dimension regardless of the distinct manner in which they are instantiated. An intermediate level was assessed by introducing novel instances of the relevant dimension; for example, training identity responding in which color is the relevant dimension and transfer testing with novel colors instantiated in the familiar training shapes. If an organism maintains discriminative responding when the instances of the relevant dimension are made novel, then the organism has demonstrated a conceptual understanding that the identity relation applies to all members of the dimension when they are instantiated in the same way. Finally, the highest level corresponded to introducing novel instances of the relevant and irrelevant dimensions in conjunction, with maintenance of discriminative responding indicating a conceptual understanding of identity for members within the dimension regardless of how they are distinctly instantiated.

2.2.3.1 Identity-Nonidentity about Color and Shape in Apes

There is only a single study addressing whole-nonidentity responding and concept formation in apes. The study showed that a language-trained chimpanzee formed a concept about shape at the intermediate level, but only after additional training with a larger set of exemplars (Tomonaga, 1995). First, a chimpanzee called Ai was trained to select the odd shaped white target stimulus from a set of identically shaped, white distracter stimuli to a criterion level of accuracy with 12 target-distracter oddity problems. Then, interspersed within baseline trials of the training problems were nonreinforced probe trials in which the shape of the target and distracter stimuli was novel, but the subject's accuracy was statistically no different from chance for the novel problems. Full transfer of learning occurred only when nonreinforced and reinforced probe trials for which the shape of the target and distracter stimuli was novel were interspersed within baseline trials of the old novel and old semi-novel transfer test problems.

With respect to investigations of part-identity and -nonidentity responding and concept formation, there is also only a single study in apes. The study showed that a group of chimpanzees formed a concept about color (or shape) as a constituent property of objects at the highest level; notably, the same could not be said of a group of orangutans (King, 1973).[4] Specifically, using a paired S/D object discrimination task, half of the subjects (the shape-same-color-different group) were trained to select the stimulus pair that had identical shapes and different colors with the incorrect pair having different shapes and identical colors. The other half of subjects (the color-same-shape-different

---

[4] By indexing concept learning with all 102 differentially reinforced transfer trials, the author concluded that the orangutans demonstrated concept formation, which is opposite to my interpretation.

group) were trained to select the stimulus pair that had identical colors and different shapes with the S- pair having different colors and identical shapes. For example, it was correct for the shape-same-color-different group to select the pair of triangles that were colored differently and the color-same-shape-different group to select the pair of red stimuli that were shaped differently when presented with a red triangle and green triangle as one pair and a red C-shape and red square as the other pair.

To test for transfer of learning, subjects were given problems constructed from three novel shapes and three novel colors. The accuracy of the orangutans was statistically at chance and improved to above chance levels with additional trials, which suggests associative learning. On the other hand, the accuracy of the chimpanzees was statistically above chance and did not improve with additional trials, which suggests partial concept formation. Furthermore, additional test trials for which the correct stimulus pair was identical in both color and shape and the incorrect stimulus pair different in color (e.g., red triangle and red triangle vs. red square and white C-shape), shape (e.g., red triangle and red triangle vs. white triangle and white C-shape) or both color and shape (e.g., red triangle and red triangle vs. red triangle and white C-shape) showed that the color-same-shape-different group chose primarily according to color identity, not shape nonidentity and that the shape-same-color-different group chose primarily according to color nonidentity, not shape identity.

2.2.3.2 Identity-Nonidentity about Color and Shape in Monkeys

Monkeys more often than not form whole-identity and -nonidentity concepts about shape. A concept of whole-nonidentity about shape formed at the intermediate level after a group of eight monkeys learned to concurrently judge color and shape whole-nonidentity (Meyer & Harlow, 1949). Specifically, subjects were trained to select the odd shaped white stimulus from two other identically shaped white stimuli (e.g., white trapezoid:white triangle:white triangle) trained to a criterion level of accuracy; thus, the odd stimulus was odd for no dimension other than shape.[5] As a group, accuracy was statistically above chance for the first 24 differentially reinforced trials of three sets that contained novel white shapes. Additionally, a concept about shape whole-identity formed at the intermediate level in two capuchin monkeys after concurrent color and shape whole-identity matching (Barros, Galvão, & McIlvane, 2002) and in three tufted capuchin monkeys after shape whole-identity matching (Truppa et al., 2010).

On the other hand, shape whole-identity responding did not lead to the formation of a concept of shape whole-identity at the intermediate level in three tufted capuchin monkeys (Truppa et al., 2010) and two Japanese monkeys (Kojima, 1979, 1982). For the aforementioned two instances that concepts failed to form, however, additional training with a larger set of exemplars resulted in concept formation. Specifically, additional shape whole-identity matching with a larger set of exemplars resulted in the formation of a concept of shape whole-identity at the highest level in six capuchin monkeys even though in the earlier experiment three subjects failed to form a shape whole-identity concept (Truppa et al., 2010). Also, a concept of whole-identity about shape did not form at the intermediate level until two Japanese monkeys received additional shape whole-identity training with a larger set of exemplars (Kojima, 1979, 1982).

---

[5] A semicolon is used to separate stimuli within oddity problems and the odd stimulus is indicated first.

With respect to color whole-identity and -nonidentity concept formation, it is clear that monkeys sometimes succeed in this regard, but they fail equally as often. At the intermediate level, a group of eight monkeys formed a whole-nonidentity concept after they learned to concurrently respond to color and shape whole-nonidentity (Meyer & Harlow, 1949) and two capuchin monkeys formed a whole-identity concept after they learned to concurrently respond to color and shape whole-identity (Barros, Galvão, & McIlvane, 2002). At the lowest level, four rhesus monkeys (Jackson & Pegram, 1970) and one Japanese monkey (Fujita, 1983b) formed a whole-identity concept and four Japanese monkeys formed a whole-identity and -nonidentity concept even though one subject required additional training with a larger set of stimuli before concept learning occurred (Fujita, 1983b). On the other hand, at the lowest level, six Japanese monkeys failed to show behavior indicative of a concept of whole-identity about color (Fujita, 1982, 1983a, 1983b) and four Japanese monkeys failed to show behavior indicative of color whole-identity and -nonidentity concept formation (Fujita, 1983a, 1983b).

Investigations about identity-nonidentity concept formation about color and shape as constituent properties of objects are represented in a limited capacity within the literature as more emphasis is placed on concept formation across objects globally. Although the empirical evidence is limited, monkeys form part-identity concepts about color and shape. First, concurrent responding to color and shape part-nonidentity during oddity tasks led to the formation of a concept about shape and color part-nonidentity at the lowest level in two rhesus monkeys and a concept about color part-nonidentity at the intermediate level in one of the two rhesus monkeys (Young & Harlow, 1943).[6] Specifically, both subjects continued to select (a) the odd shaped stimulus statistically more often than chance for two sets of novel colored, familiar shaped stimuli and (b) the odd colored stimulus statistically more often than chance for two sets of novel shaped, familiar colored stimuli. Further, one subject continued to select the odd colored stimulus statistically more often than chance for two sets of novel colored, familiar shaped stimuli, which illustrates concept formation at the intermediate level. Second, a part-identity concept about color formed at the lowest level after color part-identity responding in two rhesus monkeys even though one monkey required additional color part-identity matching with a larger set of exemplars before concept formation occurred (Weinstein, 1945).

2.2.3.3 Identity-Nonidentity about Color and Shape in Infants

Only a few empirical reports investigate concept formation using operant and associative learning methods in infants. The findings indicate that nonverbal infants form whole-identity concepts about shape and part-identity and -nonidentity concepts about shape and color. The major methodological limitation of these experiments is that they do not examine the data for individual differences in performance so it is possible that many individuals within the group failed to form concepts. Identifying individual differences in performance could enhance the generalizability of results and show that identity judgments can establish a range of stimulus control topographies, which permit

[6] My conclusion differs from the authors' who based their conclusion of successful generalization of color and shape oddity judgments on transfer tests that involved previously used, differentially reinforced transfer stimuli, transfer stimuli that may not have been discriminably different from the training stimuli, and nonsimilarity (instead of nonidentity) based judgments.

individual acquisition and promote or retard generalization (Brannon, Cantlon, & Terrace, 2006; Galvão et al., 2005; Judge, Evans, & Vyas, 2005).

Using the visual anticipation paradigm, the responses of nine 5- to 7-month-olds indicated the formation of a whole-identity concept about shape (McMurray & Aslin, 2004). During the training phase, participants were shown shapes in one color that appeared in the center of the screen before disappearing from the screen (e.g., a yellow cross and yellow square). Visual reinforcement then appeared on the left side of the screen if the previous stimulus was the square and on the right side of the screen if the previous stimulus was the cross. Each participant learned the relation between shape and the spatial location of reinforcement because they looked at the correct side of the screen for a longer period than the incorrect side of the screen statistically more often than what chance would predict. During the transfer testing phase, nonreinforced test trials of the familiar shapes in novel colors we shown (e.g., red square and orange cross). Correct anticipatory eye movements did not differ statistically between nonreinforced trials of the training stimuli and the novel colored, familiar shaped test stimuli; thus, infants recognized that shape was invariant in the face of changes to its color and their anticipatory eye movements followed the pattern learned during training.

Second, using the conjugate reinforcement paradigm, a part-identity concept about shape formed at the highest level in a group of 2.5 to 4 month old infants (Hayne, Rovee-Collier, & Perris, 1987). Training mobiles consisted of wooden blocks with the same colored shape on each block (e.g., blocks with red As). The shape on the blocks remained unchanged during training, while the color varied (e.g., red As, blue As, and green As). During transfer testing, participants were shown mobile blocks with the familiar or a novel shape in one novel color (e.g., black As or black 2s). If participants learned the relation between the shape on the mobile blocks and activation of the mobile, then they should continue kicking when the familiar shape was shown in a novel color and discontinue kicking otherwise. During the transfer test, the rate of kicking was statistically higher than the baseline rate of kicking for the familiar shape in the novel color, which shows that infants generalized their responses to members of the class (e.g., the shape on blocks), and was statistically no different from baseline for the novel shape in the novel color, which shows that they did not generalize their responses to nonmembers of the class. The results showed that the infants formed a part-identity concept about shape as a constituent property of the mobiles. Finally, an additional experiment showed conceptualization of identity about both color and shape as the constituent properties of objects at the lowest level in two groups of six-month-olds, but not in a group of three-month-olds using the conjugate reinforcement paradigm (Bhatt, Wilk, Hill, & Rovee-Collier, 2004).

2.2.3.4 Discussion

The levels of concept formation proposed in this dissertation may be analogous to the levels of relational knowledge that were proposed by Smith (1984) to underlie the development of dimensional comparisons. The first level of understanding relations involves knowledge that a particular attribute (e.g., red) can be instantiated in a variety of distinct objects. The second level involves knowledge that there are qualitatively distinct kinds of attributes (e.g., red and blue are attributes of the same kind) that are different from other kinds, in other words, knowledge that a particular dimension exists (e.g., the color dimension). Young children are able to learn relations at the first level, whereas, only older children are generally able to learn relations at the second level. With that said, the lowest level of concept formation proposed in this dissertation may be akin to the first level of relational knowledge, the intermediate level akin to the second level of relational knowledge, and the highest level an integration of the first and second levels of relational knowledge. If this is the case, then it should be easier to form concepts at the lowest level and hardest to form concepts at the highest level. The just reviewed literature does not disentangle these relationships so it remains a matter for future research.

The differentiation hypothesis about conceptual development proposes that there is a shift during infancy from the discovery of identity relations based on the global or holistic aspect of objects (i.e., across all dimensions at once) to those that are articulated along the different dimensions of objects. What this means is that infants first form broad, global concepts and categories (i.e., whole-identity) for which the members are related to one another based on overall similarity, but as infants age and gain experience with the world, they progress towards forming concepts for which the members are related to each other by the possession of a common property (i.e., part-identity). Thus, concept development is assumed to begin at a concrete level (e.g., these things all look alike) and become more abstract (e.g., these things are all the same kind of thing) as children learn to generalize across more and more varied instances (Burns, 1992; Kemler, 1983; Smith, 1984; Smith, 1993; Smith & Heise, 1992). Researchers posit that hearing different words consistently applied to two objects stimulates nonverbal infants to pay attention to the perceptual variants between those two objects that previously were interpreted as minor (Mandler, 2004).

So what does the empirical evidence detailed in this dissertation reveal in this regard? Although circumstantial in nature, it is possible that concepts about part-identity do not form in young infants as the empirical evidence indicated that conceptualization of identity about color and shape as the constituent properties of objects occurred in 2.5- to 4-month-olds (Hayne, Rovee-Collier, & Perris, 1987) and 6-month-olds (Bhatt et al., 2004), but not 3-month-olds (Bhatt et al., 2004). In apes and monkeys, the differentiation hypothesis may translate into a difference in the ease of forming part- and whole-identity concepts; clearly, more research is necessary to assess the differentiation hypothesis.

Another unanswered question that arises from the empirical literature concerns whether it is easiest to form concepts about color or concepts about shape. There is evidence in apes indicating that it is harder to learn how to match according to shape part-identity than color part-identity, but there is also evidence indicating that shape and color part-identity matching is equally difficult. Specifically, five chimpanzees were more accurate for color than shape part-identity problems (average 11% difference) when blocks of color and shape part-identity trials were presented using a two-choice conditional MTS procedure (e.g., green triangle → green square, not red triangle when presented on a wooden platform, but vice versa when presented on a metal platform) (Nissen, Blum, & Blum, 1949). Additionally, when color and shape part-identity trials were intermixed to require concurrent matching for three of the aforementioned chimpanzees, they were still more accurate for color than shape part-identity problems (average 14% difference). On the other hand, shape and color part-identity matching accuracy was not statistically different for four orangutans and one gorilla (average 5% difference) when irrelevant dimensions were cue-ambiguous (e.g., green circle → green cross, not red square; green triangle → yellow triangle, not blue square) (Vonk, 2003). Again, more research is necessary to determine if the same pattern holds for concept formation.

CHAPTER 3

EXPERIMENTAL INVESTIGATION OF COLOR AND SHAPE IDENTITY-NONIDENTITY RESPONDING AND CONCEPT FORMATION

The objective of Experiment 1 was to establish concurrent whole-identity responding with a set of stimuli that differed in only one regard, their color, using a novel conditional discrimination procedure. The hypothesis was that subjects would learn to respond to color whole-identity and -nonidentity; further, I hypothesized that there would be accuracy and response time differences between color whole-identity and -nonidentity judgments.

The objective of Experiment 2 was to establish concurrent color and shape part-identity responding with a small set of colors and shapes. The hypothesis was that subjects would learn to respond concurrently to both color and shape part-identity; further, I hypothesized that there would be accuracy and response time differences between color and shape part-identity judgments.

Finally, the objective of Experiment 3 was to establish concurrent color and shape part- and whole-identity responding with a small set of familiar colors and shapes and then test for concept formation at the highest level by introducing problems in which both the color and shape of stimuli was novel. It was predicted that it would be more difficult to form part-identity concepts about color and shape than whole-identity concepts about color and shape.

3.1 Method

3.1.1 Subjects and Housing

The subjects were two Sumatran orangutans (*Pongo pygmaeus abelii*), Madu and Junior (also called Bernas), that were housed at Zoo Atlanta. Madu was a captive-born, hand-reared adult female aged 26.5 years at the start of the study. Junior was a captive-born, foster-reared (by Madu), developing adolescent male aged 6.7 years at the start of the study. Madu had an extensive experimental history of participating in cognition and learning tasks that began with her involvement in a joystick-controlled computerized route and detour task when she was around six years old (Menzel & Menzel, 2007). Junior was experimentally naïve having never participated in any form of cognition and learning task. An additional three orangutans, two adult males and one adult female (Allen, Jantan, and Biji), were excluded from further participation in the study because they ceased to respond during the first or third phase of shaping or they exhibited a positional response bias during Experiment 1A.

Subjects were not food- or water-deprived during the study. The two subjects were socially housed together with a dominant adult male and an infant orangutan in large indoor and outdoor enclosures. The caging of the indoor enclosures was made of a plain weave wire mesh (about 9.5 mm wire diameter with 25 mm wire openings).

3.1.2 Apparatus and Materials

A touchscreen frame (70.9 x 39.9 cm active area) held LEDs that created a grid of infrared lights such that x- and y-coordinates were identified when a touch obstructed one or more beam.[7] A flatscreen LCD television (69.7 x 39.2 cm viewing area) sat behind the touchscreen to display the visual stimuli and sound the auditory stimuli. The television's resolution was set to 1366 x 768 pixels; thus, each pixel (dot pitch) measured 0.51 mm x

---

[7] All measurements are length by height by width.

0.51 mm and there were 49.75 pixels per inch (PPI). Behind the touchscreen frame and in front of the television sat a clear acrylic shield (72 x 42 cm) that prevented subjects from contacting the television, but allowed subjects to view stimuli displayed on the television. The touchscreen frame, acrylic shield, and television are collectively called the touchscreen. The touchscreen was connected to the experimenter's (U.A.) laptop computer that ran the stimulus presentation software (SuperLab 4.0, Cedrus Corporation, San Pedro, CA) that delivered trials of the visual and auditory stimuli and recorded subject responses and response times. The visual stimuli presented by the stimulus program were generated by a custom computer program.

Food was dispensed manually via a PVC pipe that was angled downward such that food traveled from the upper to the lower opening and exited inside the enclosure of participants. Yogurt served as the food reinforcer during the shaping phase and frozen grapes of approximately the same size (pieces cut into 1/2 to 1/8) served as the food reinforcers during the part and phases of the experiment s.

The touchscreen was secured on the lower shelf of a moveable cart about 15 cm off the ground and faced towards subjects. The experimenter's laptop rested on the upper shelf of the cart (i.e., the laptop station) about 55 cm off the ground and faced away from subjects. An acrylic sheet (1.2 x 1.2 m) with a cutout (70 x 39 cm) to allow touches to the touchscreen was attached to the face of the cart. The acrylic sheet limited physical contact to just the touchscreen and obscured the face and body of the experimenter from subjects.

3.1.3 Visual and Auditory Stimuli

Visual stimuli consisted of seven shapes: square, rectangle, pentagon, circle, rhombus (diamond), trapezoid, and hexagon. The shapes were shown in seven colors: red (RGB: 255, 000,000), yellow (RGB: 255, 255, 000), brown (RGB: 153, 102, 000), pink (RGB: 255, 000, 135), purple (RGB: 085, 000, 085), orange (RGB: 055, 102, 000), and grey (RGB: 128, 128, 128). Additionally, the square and rectangle shapes were shown in a black and white pattern.

The sample-touched tone was one 300-ms sound (suction.wav) from Microsoft Office 2003. The correct-response tone was one 500-ms high (1,000 Hz) tone and the incorrect-response tone was one 1,000-ms low (100 Hz) tone, both of which were created using a freeware sound generator program.

3.1.4 General Procedure

Subjects were isolated from the rest of their social group or just isolated from dominate members of their social group when participating in the study. Specifically, Junior was always isolated from Madu and the dominant adult male and for the majority of time isolated from the infant orangutan. Madu was always isolated from the dominant adult male, sometimes isolated from Junior, and almost never isolated from the infant orangutan. Subjects had visual and auditory contact with the other members of their social group while they participated in the study.

Sessions started with the experimenter positioning the apparatus flush against the cage mesh of the indoor enclosures so that subjects were able to reach through the mesh and touch the touchscreen. Then, the experimenter sat behind the apparatus at the laptop station to start trials, deliver food reinforcers, and monitor the presentation of stimuli. For the majority of trials within a session, the experimenter was blind, a situation that was created by lowering or looking away from the laptop screen; thus, food reinforcers were delivered or not delivered based on the tone that sounded. For a small number of trials within a session, the experimenter was not blind, but instead monitored the delivery of trials on the laptop display screen to ensure that trials were presented free of experimenter error and computer malfunction.

Except for some of the phases of shaping, sample stimuli were displayed centered in the top half of the screen and comparison stimuli were displayed in the bottom half of the screen side-by-side and separated by at least 100 pixels in three possible spatial positions: left, right, or center. A trial started with the sample stimulus displayed on the screen. A touch to the sample stimulus sounded the sample-touched tone and then the sample stimulus became unresponsive to touches, but it remained on the screen. Following a touch to the sample, the comparison stimuli appeared on the screen. Correct responses to the comparison stimuli were followed by the correct-response tone and the experimenter delivering food reinforcers. Incorrect responses were followed by the incorrect-response tone. Correct and incorrect responses were both followed by an intertrial interval (ITI). The ITI involved showing a blank (black) screen for some amount of time before the next trial was presented during which time responses were not registered by the touchscreen. Figure 3.1.1 shows the trial sequence.[8]

Reinforcement was delivered on a continuous reinforcement (CRF) schedule. Response time was measured from the onset time of the comparison stimuli to a subject's touch. One hundred trials were presented within a session (except during Shaping and

[8] In all figures, letters are shown within stimuli to aid readability. Additionally, except for Figure 3.1.1, figures depict problems following a touch to the sample (i.e., sample and comparison stimuli are displayed simultaneously).

Experiment 1A, Phase 2) unless a session was terminated because of husbandry procedures, subject unresponsiveness, equipment malfunction, or experimenter error. Subjects advanced to a different experiment part (A, B, C, etc.) or phase (1, 2, 3, etc.) when they meet a specified performance criteria.

3.1.4.1 Shaping

Shaping occurred between April 6, 2009 and May 22, 2009. One 2- to 150-trial session occurred one to four days per week for a total of 20 days. Only Junior participated in shaping as Madu was already experienced with touchscreen-controlled two- and three-choice MTS tasks. There were four phases of shaping and each phase utilized grey squares or rectangles as the stimuli.

The first phase of shaping established touching of the touchscreen by presenting errorless trials. Before activating the touchscreen, the experimenter applied yogurt to the screen locations that corresponded to where the sample and two comparison stimuli would appear. Trials started with the sample stimulus and two comparison stimuli simultaneously displayed. Touching the sample stimulus sounded the sample-touched tone and touching either of the two comparison stimuli sounded the correct-response tone. For this phase only, reinforcement was delivered after both the sample- and the correct- response tone regardless of whether or not the subject was focused on consuming yogurt. The sample-touched and correct-response tones were followed by a 0.5-s ITI. The sample stimulus spanned most of the top half of the screen (1300 x 375 pixels). The two comparison stimuli were the same size (600 x 375 pixels), thus, collectively they spanned most of the bottom half of the screen.

The second phase of shaping established the touching of stimuli regardless of where they were displayed on the touchscreen using a fading procedure. Trials started by displaying a target stimulus. Touching the target stimulus sounded the correct-response tone, which was then followed by the delivery of reinforcement and a 2-s ITI. Initially, the target stimulus was a gray rectangle (1200 x 750 pixels) that spanned most of the screen and its spatial position randomly moved between four locations (upper right, upper left, lower right, and lower left) across trials. Over the course of shaping sessions, the size of the target stimulus was reduced to a medium sized square (500 x 500 pixels) and the possible spatial positions increased to nine screen locations (upper right, middle right, lower right, upper center, middle center, lower center, upper left, middle left, and lower left).

The third phase of shaping involved chaining the appropriate response sequence (i.e., touch the sample, then touch a comparison stimulus) when the sample and one comparison stimulus was displayed. Trials proceeded as described in the General Procedure section except that only one comparison stimulus was utilized with its left or right spatial position randomly determined. A touch to the comparison stimulus sounded the correct-response tone, which was followed by the delivery of reinforcement and a 2-s ITI. The sample and comparison stimulus were squares of the same size (350 x 325 pixels).

The fourth phase of shaping involved chaining the appropriate response sequence when the sample and two comparison stimuli were displayed. Trials proceeded as described in the General Procedure except that one of the two comparison stimuli was responsive to touches while the other was unresponsive to touches. A touch to the responsive comparison stimulus sounded the correct-response tone, which was followed by the delivery of reinforcement and a 2-s ITI. The sample and comparison stimulus were squares of the same size (350 x 325 pixels).

3.1.4.2 Experiment 1A: Color Whole-Identity Responding

Subjects participated in two- and three-choice matching-to-sample (MTS) tasks for one to two sessions per day: Junior for 42 days between August 3, 2009 and November 20, 2009 and Madu for 5 days between September 8 and 15, 2009 (Phase 1 only). The purpose of this experiment was to establish color whole-identity responding with a small set of color stimuli that were instantiated in one shape.

The sample and comparison stimuli were colored yellow (Y), red (R), and brown (B) and they were shaped as squares of the same size (350 x 325 pixels). A correct response was to touch the comparison stimulus that was the same color as the sample stimulus (i.e., touching the correct or matching comparison stimulus). An incorrect response was to touch a comparison stimulus that was a different color than the sample stimulus (i.e., touching an incorrect or a nonmatching comparison stimulus). Thus, color was the relevant dimension and shape was an irrelevant cue-constant dimension. Shape was not only cue-constant within trials, but it was also cue-constant between trials because it never varied.

Trials were separated by a 3-s ITI. The number of times that each color served as the sample stimulus was balanced within sessions and the presentation order for the sample's color was randomized within sessions. The spatial position of the comparison stimuli was randomized across trials within a session. Unless otherwise noted, two-choice MTS tasks utilized two spatial positions for the comparison stimuli (e.g., the left and right positions) and three-choice MTS tasks utilized three spatial positions for the comparison stimuli (left, right, and center positions). The performance criteria were 80% or more correct responses for three consecutive sessions with a minimum of five completed sessions.

3.1.4.2.1 Three-choice MTS (Phase 1)

Three comparison stimuli were used; thus, three problems were presented: B → B, not R and Y; R → R, not B and Y; and Y → Y, not B and R. Figure 3.1.1 shows an example problem and trial sequence. Subjects advanced to Experiment 1B when they meet the performance criteria; otherwise, they advanced to the Phase 2 after completing 1,500 trials.

3.1.4.2.2 Three-choice MTS with Correction Procedure (Phase 2)

This purpose of this phase was to establish color whole-identity responding using a correction procedure. Trials were identical to Phase 1 except for the following details. A correction procedure was utilized such that trials to which a subject responded incorrectly were represented until they made a correct response. During correction trials, the left, right, and center spatial position of comparison stimuli was re-randomized. One session occurred per day during which the subject received from 71 to 222 noncorrection and correction trials per day. Figure 3.1.1 shows an example problem and trial sequence. The subject advanced to Phase 3 when he did not meet the performance criteria after completing approximately 1,300 trials.

3.1.4.2.3 Two-choice Three-position MTS (Phase 3)

This purpose of this phase was to establish color whole-identity responding by changing the format of the task to two-choice three-position. Trials were identical to Phase 1 except for the following details. Only two comparison stimuli were utilized (i.e., the left, right, or center spatial position remained empty); thus, six problems were presented: B → B, not R or Y; R → R, not B or Y; and Y → Y, not B or R. Figure 3.1.1 shows an example problem and trial sequence. The subject advanced to Phase 4 when he met the performance criteria.

3.1.4.2.4 Two-choice MTS (Phase 4)

The purpose of this phase was to ensure stable color-whole identity responding when only two spatial positions for the comparison stimuli were used. Trials were identical to Phase 1 except for the following details. The spatial position of stimuli was restricted to only the left or right position (i.e., the center spatial position was not used). Figure 3.1.1 shows an example problem and trial sequence. The subject advanced to Experiment 1B, Phase 2 when he met the performance criteria.

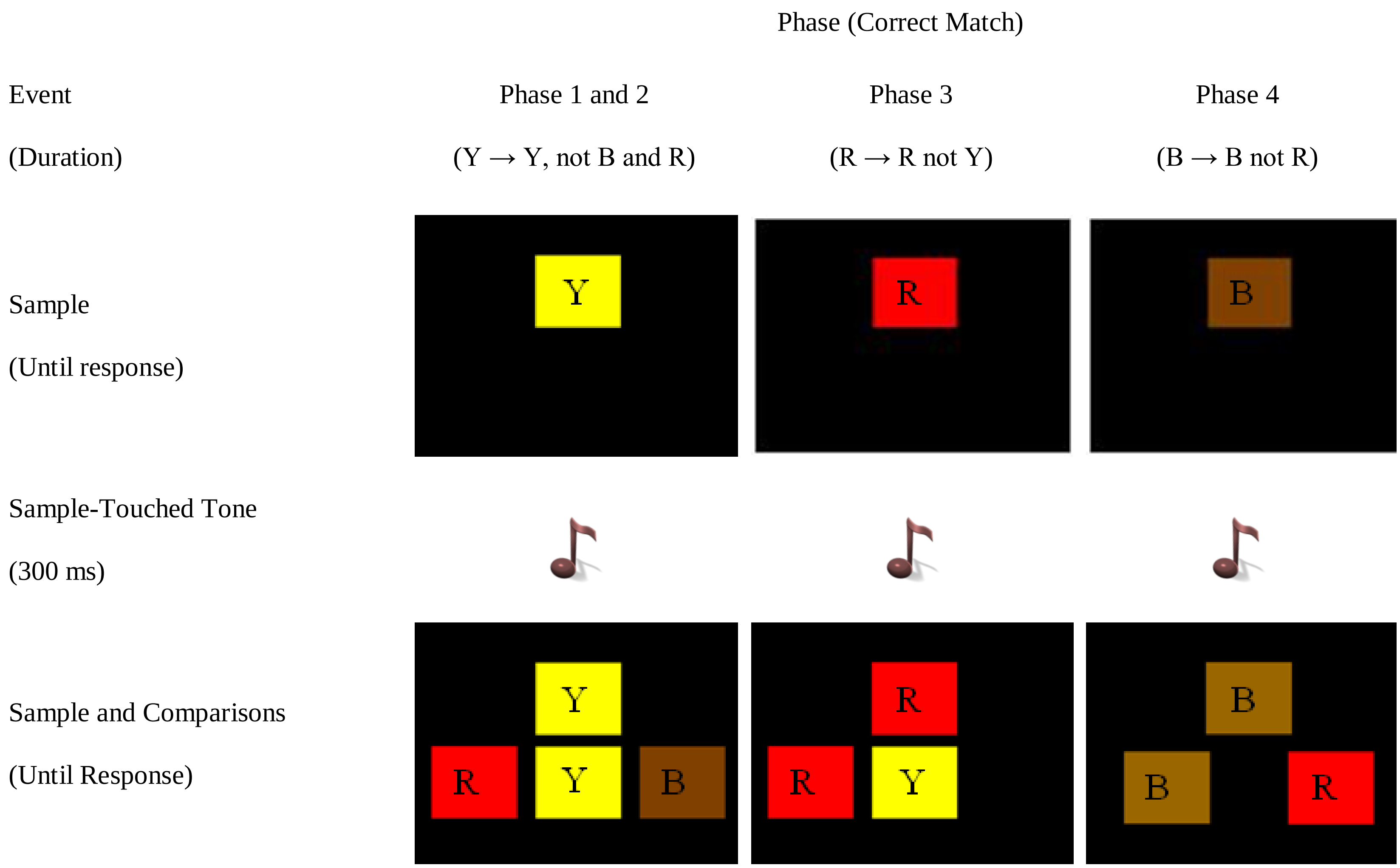


Figure 3.1.1. Trial sequence and example problems for color whole-identity responding in Experiment 1A.

Correct-Response Tone (500 ms)

or

Incorrect-Response Tone (1,000 ms)

Reinforcement

or

Nonreinforcement

ITI Blank Screen

(3,000 ms)

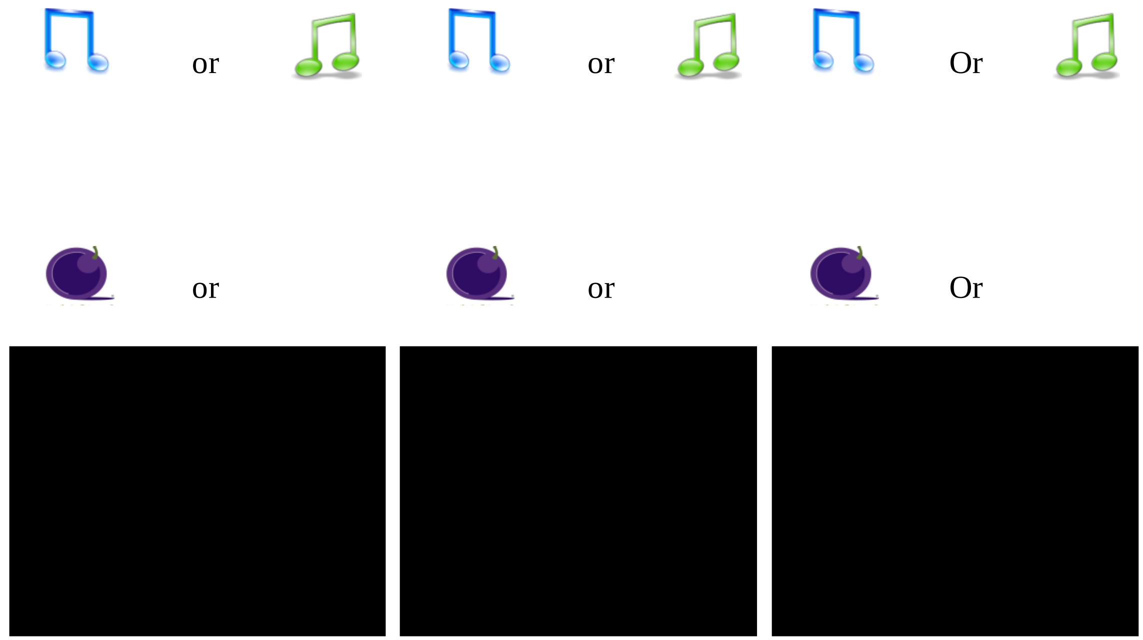


Figure 3.1.1 Continued.

3.1.4.3 Experiment 1B: Concurrent Color Whole-Identity and -Nonidentity Responding

Subjects participated in two-choice modified MTS tasks for one to two sessions per day: Junior for 7 days between November 23, 2009 and December 7, 2009 (Phase 2 only) and Madu for 10 days between November 3 and 17, 2009.[9] The purpose of this part of the experiment was to establish concurrent whole-identity and -nonidentity responding to a small set of color stimuli that were instantiated in three shapes.

The procedures of Experiment 1A were replicated to require whole-identity responding. Additionally, a black and white pattern was added to the set comparison stimuli colors. The patterned comparison stimulus served as the none-of-the-above comparison stimulus (NOTA) to allow for whole-nonidentity judgments. Thus, concurrent identity and nonidentity responding was required. The performance criteria was 80% or more correct responses for three consecutive sessions for both identity and nonidentity problems with a minimum of five sessions completed. The number of times the NOTA stimulus served as a comparison stimulus was balanced within sessions.

There were nine whole-identity problems for which it was correct to select the comparison stimulus that matched the color of the sample stimulus. Of these, NOTA was absent for one comparison stimulus pair type (B → B, not R or Y; R → R, not B or Y; and Y → Y, not R or Y) and NOTA was the incorrect comparison for one comparison stimulus pair type (B → B, not NOTA; R → R, not NOTA; and Y → Y, not NOTA).

---

[9] Before participating in this experiment, Madu completed two- and three-choice MTS tasks with a different NOTA stimulus: a black and white patterned X-shape. Madu failed to reach the performance criteria after 1,500 trials, but it was likely that she was matching based on shape; that is, avoiding the X-shaped NOTA comparison because it was shaped differently from the square shaped color comparisons. As such, the shape of the NOTA stimulus was changed to a square.

There were also six nonidentity problems for which the sample stimulus was presented with a nonmatching comparison and the NOTA comparison stimulus and it was correct to select the NOTA comparison stimulus. These problems were called the NOTA correct comparison stimulus pair type: B → NOTA, not R or Y; R → NOTA, not B or Y; and Y → NOTA, not B or R.

3.1.4.3.1 Two-choice Three-position Modified MTS (Phase 1)

The purpose of this phase was to introduce concurrent identity and nonidentity responding in a familiar format (i.e., three spatial positions for the comparison stimuli) before exposure to it in an unfamiliar format for Madu who had not yet experienced two-choice MTS tasks. Two comparison stimuli were simultaneously presented in three spatial positions (i.e., the left, right, or center position remained empty). Figure 3.1.2 shows an example identity problem in which NOTA is the incorrect comparison and an example nonidentity problem in which NOTA is the correct comparison. Refer to Figure 3.1.1 for illustrations of color whole-identity problems in which NOTA is absent. The subject advanced to Phase 2 when she met the performance criteria.

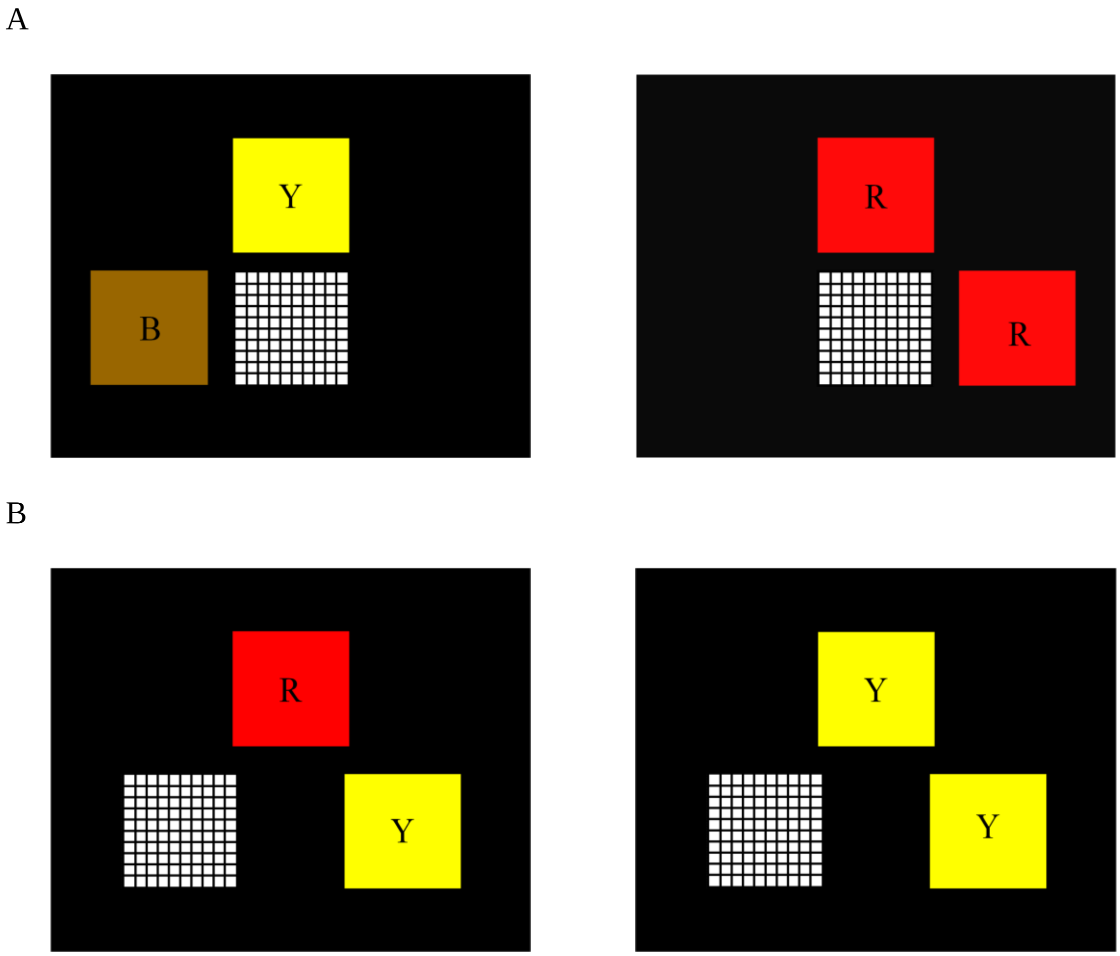


Figure 3.1.2. Example problems during concurrent color whole-identity and -nonidentity responding in Experiment 1B. (A): On the left, a color nonidentity problem in which NOTA is the correct comparison (Y → NOTA, not B) and on the right, a color identity problem in which NOTA is the incorrect comparison (R → R, not NOTA) for the two-choice three-position MTS tasks (Phase 1). (B): On the left, a color nonidentity problem in which NOTA is the correct comparison (R → NOTA, not Y) and on the right, a color identity problem in which NOTA is the incorrect comparison (Y → Y, not NOTA) for the two-choice MTS tasks (Phase 2).

3.1.4.3.2 Two-choice Modified MTS (Phase 2)

The purpose of this phase was to establish concurrent identity and nonidentity responding using a two-choice format. Trials were identical to Phase 1 except that the spatial position of comparison stimuli was restricted to only the left or right spatial position (i.e., the center position was not used). Figure 3.1.2 shows an example identity problem in which NOTA is the incorrect comparison and an example nonidentity problem in which NOTA is the correct comparison. Refer to Figure 3.1.1 for illustrations of color whole-identity problems in which NOTA is absent. Subjects finished Experiment 1 when they met the performance criteria.

3.1.4.4 Experiment 2A: Concurrent Color and Shape Part-Identity Responding

Subjects participated in two-choice MTS tasks for one to two sessions per day: Junior for 20 days between June 12, 2010 and October 10, 2010 and Madu for 18 days between June 12, 2010 and October 9, 2010. The purpose of this part of the experiment was to establish concurrent color and shape part-identity responding to a small set of color and shapes.

For color part-identity problems, color was the relevant dimension (i.e., correlated with differential reinforcement) and shape was the irrelevant cue-ambiguous dimension (i.e., shape differed among all stimuli within a trial); thus, the correct response was to touch the comparison stimulus that was the same color as the sample stimulus and an incorrect response was to touch the comparison stimulus that was a different color than the sample stimulus. For shape part-identity problems, shape was the relevant dimension and color was the irrelevant cue-ambiguous dimension; thus, the correct response was to touch the comparison stimulus that was the same shape as the sample stimulus and the incorrect response was to touch the comparison stimulus that was a different shape than the sample stimulus.

Sample and comparison stimuli were colored yellow, red, and brown. Sample and comparison stimuli were shaped as rectangles (Rect), pentagons (Pent), and circles (Circ) of approximately the same area (90,100 pixels, 89,917 pixels, and 89,529 pixels, respectively). There were six color part-identity problems: B → B, not R or Y; R → R, not B or Y; and Y → Y, not B or R. There were six shape part-identity problems: Circ → Circ, not Pent or Rect; Pent → Pent, not Circ or Rect; and Rect → Rect, not Circ or Rect. Figure 3.1.3 shows example color and shape part-identity problems.

Figure 3.1.3. Example problems for color and shape part-identity responding in Experiment 2A. The left panel shows a color part-identity problem (B → B, not R). The right panel shows a shape part-identity problem (Pent → Pent, not Circ).

Half of the trials in a session were color part-identity problems and the other half of trials were shape part-identity problems such that concurrent color and shape identity responding was required. The number of times that each color and shape served as the sample stimulus was balanced within sessions and the presentation order for the sample's color and shape was randomized within sessions. Color and shape part-identity problems were randomly interspersed together within sessions. The spatial position of stimuli was restricted to only the left or right position (i.e., the center spatial position was not used) and randomized across sessions. Unless otherwise noted, two-choice MTS tasks utilized two spatial positions for the comparison stimuli (the left and right positions). Trials were separated by a 2-s ITI. If after completing 3,000 trials subject accuracy was not at least 80% correct for each type of identity problem (i.e., for both color part-identity and shape part-identity problems) for three consecutive sessions, they advanced to Experiment 2B.

3.1.4.5 Experiment 2B: Shape Part- and Whole-Identity Responding

Subjects participated in two-choice MTS tasks for one to four sessions per day: Junior for 23 days between October 11, 2010 and November 5, 2010 and again for 6 days between November 5 and 10, 2010 and Madu for 3 days between October 12 and 14, 2010 (Phase 1 only). The purpose of this part of the experiment was to establish in sequence shape part- and whole-identity responding after a failure to establish concurrent color and shape part-identity responding.

3.1.4.5.1 Shape Part-Identity (Phase 1)

The purpose of this phase was to establish shape part-identity matching. All trials within a session were shape part-identity problems like those in Experiment 2A. Subjects advanced to Experiment 2C when they met the performance criteria; otherwise, they advanced to Phase 2 after completing 1,500 trials.

3.1.4.5.2 Shape Whole-Identity (Phase 2)

This purpose of this phase was to establish shape whole-identity matching. All trials within a session were shape whole-identity problems. That is, shape was the relevant dimension and color was the irrelevant cue-constant dimension (i.e., color was identical among all stimuli in a trial); thus, the correct response was to touch the comparison stimulus that was the same shape as the sample stimulus and the incorrect response was to touch the comparison stimulus that was a different shape than the sample stimulus.

The number of times that each shape served as the sample stimulus was balanced within sessions and the trial presentation order for the sample's shape was randomized within sessions. There were six shape whole-identity problems: Circ → Circ, not Pent or Rect; Pent → Pent, not Circ or Rect; and Rect → Rect, not Circ or Rect. Figure 3.1.4 shows an example shape whole-identity problem. Other than the aforementioned, the procedures of Experiment 2A were replicated. The subject advanced to Phase 3 when he met the performance criteria.

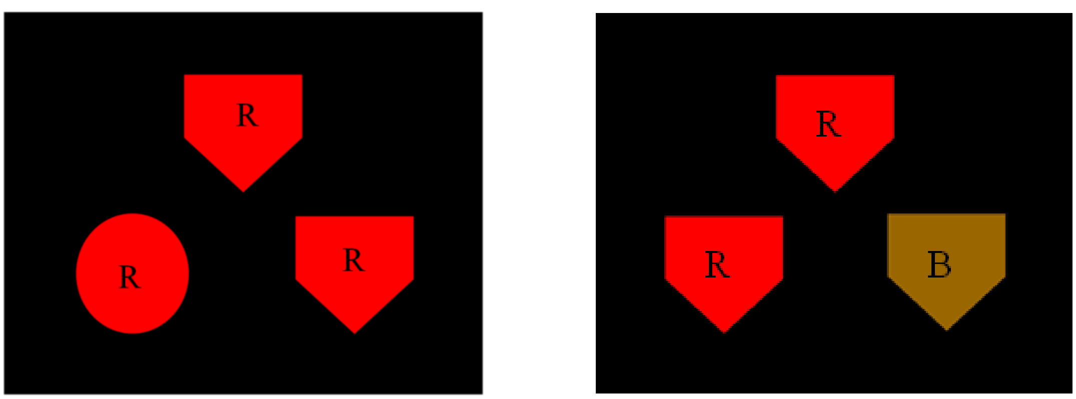


Figure 3.1.4. Example problems for shape whole-identity responding in Experiment 2B and color whole-identity responding in Experiment 2D. The left panel shows a shape whole-identity problem (Pent → Pent, not Circ). The right panel shows a color whole-identity problem (R → R, not B).

3.1.4.5.3 Shape Part-Identity (Phase 3)

This phase was an exact replication of Phase 1 with the purpose of establishing shape part-identity matching. The subject advanced to Experiment 2C when he met the performance criteria.

<u>3.1.4.6 Experiment 2C: Concurrent Color and Shape Part-Identity Responding</u>

This part of the experiment replicated the two-choice MTS tasks of Experiment 2A for two to three sessions per day: Junior participated for 7 days between November 10 and 17, 2010 and Madu for 5 days between October 15 and 20, 2010. The purpose was to establish concurrent color and shape part-identity responding. Subjects advanced to Experiment 2D when they meet the performance criteria.

<u>3.1.4.7 Experiment 2D: Color and Shape Whole-Identity Responding</u>

Subjects participated in two-choice MTS tasks for two to three sessions per day: Junior for 2 days on December 21 and 22, 2010 (Phase 1 only) and Madu for 4 days between December 31, 2010 and January 3, 2011. The purpose of this part of the experiment was to require color whole-identity responding with the full set of colors and shapes (c.f. only one shape was used in color whole-identity responding in Experiment 1) and to create equivalence between subjects in the type of identity task they completed, Madu was required to complete shape whole-identity tasks.

3.1.4.7.1 Color Whole-Identity (Phase 1)

All trials within a session were color whole-identity problems so color was the relevant dimension and shape was the irrelevant cue-constant dimension. A correct response was to touch the comparison stimulus that was the same color as the sample stimulus and the incorrect response was to touch the comparison stimulus that was a different color than the sample stimulus.

The number of times that each color served as the sample stimulus was balanced within sessions and the trial presentation order for the sample's color was randomized within sessions There were six color whole-identity problems: B → B, not R or Y; R → R, not B or Y; and Y → Y, not B or R. Figure 3.1.4 shows an example color whole-identity problem. Otherwise, the procedures of Experiment 2A were implemented. If subjects completed Experiment 2B, Phase 2 at an earlier time, then they finished Experiment 2 when their accuracy was at least 80% correct for three consecutive sessions; otherwise, subjects advanced to Phase 2.

3.1.4.7.2 Shape Whole-Identity (Phase 2)

This experiment replicated Experiment 2B, Phase 2 by presenting only shape whole-identity problems to Madu. Figure 3.1.4 shows example color whole-identity and shape part-identity problems. The subject finished Experiment 2 when she met the performance criteria.

<u>3.1.4.8 Experiment 3: Concurrent Color and Shape Part- and Whole-Identity Responding and Transfer Test</u>

This experiment presented two-choice MTS tasks with the color and shape part- and whole-identity problems given in Experiment 2; thus, concurrent identity responding was required. Then, transfer of learning was assessed at the highest level by introducing new color and shape part- and whole-identity problems for which color and shape was novel. From one to nine sessions occurred per day with Junior participating for 12 days between December 27, 2010 and January 10, 2011 and Madu participating for 7 days between January 4 to 11, 2011.

Two-choice MTS tasks utilized two spatial positions for the comparison stimuli (the left and right positions) and the spatial position of comparison stimuli was randomized for every trial. Trials were separated by a 2-s ITI. When subject accuracy was at least 70% correct for each type of identity problem (i.e., for color part-identity, color whole-identity, shape-part-identity, and shape whole-identity problems) for three consecutive sessions and they completed a minimum of five sessions they advanced to the next phase.

3.1.4.8.1 Baseline (Phase 1)

Subjects were presented with color and shape part-identity problems like that described for Experiment 2A and 2C and color and shape whole-identity problems like that described for Experiment 2B and 2D. Color and shape part- and whole-identity trials were randomly interspersed together within a session. Because these trials present the familiar, trained color and shape part- and whole-identity problems they are called baseline trials.

Within each session, one-fourth of baseline trials were color part-identity problems, one-fourth color whole-identity problems, one-fourth shape part-identity problems, and one-fourth shape whole-identity problems. For color part- and whole-identity baseline trials, the number of times that each color served as the sample stimulus was balanced within sessions. For shape part- and whole-identity baseline trials, the number of times that each shape served as the sample stimulus was balanced within sessions. Figures 3.1.3 and 3.1.4 display example color and shape whole- and part-identity problems. When Junior and Madu met the performance criteria, they advanced to Phase 2 and Phase 3, respectively.

3.1.4.8.2 Baseline and Nonreinforced (36%) Baseline (Phase 2)

The purpose of this phase was to familiarize subjects to nonreinforced trials; thus, the familiar, trained color and shape part- and whole-identity problems were presented to subjects as two types of trials. Nonreinforced baseline trials were trials in which responses were differentially reinforced like in Phase 1 and nonreinforced baseline trials were trials that were not followed by the correct- or incorrect-response tone and food reinforcement regardless of the subject's response. Of the 100 trials in a session, 36% were nonreinforced baseline trials and the remaining 64% were reinforced baseline trials. Reinforced and nonreinforced baseline trials were pseudo randomly interspersed together within a session in a way that prevented more than three consecutive nonreinforced trials. When the subject reached the performance criteria, he received an additional 1,100 trials before advancing to Phase 3 to allow his accuracy with respect to each type of identity problem to stabilize.

3.1.4.8.3 Baseline and Nonreinforced (24%) Baseline (Phase 3)

The purpose of this phase was to reduce the percentage of nonreinforced baseline trials. This phase replicated Phase 2 except for the following. First, the size of the sample and comparison stimuli was reduced to prepare for the size of the novel stimuli to be used in the transfer test of Phase 4: rectangles (59,684 pixel area), pentagons (60,172 pixel area), and circles (59,773 pixel area). Second, the percentage of nonreinforced baseline trials was reduced to 24% of trials while the percentage of reinforced baseline trials constituted 76% of trials to reduce their prominence within a session. When Junior and

Madu reached the performance criteria, they received respectively one and five additional sessions before advancing to transfer testing in Phase 4 so that the total number of trials they completed was 1,000 trials.

3.1.4.8.4 Transfer Test (Phase 4)

To test for transfer of responding from the familiar problems t o novel problems, this phase replicated Phase 3 except for the following. The percentage of nonreinforced baseline trials was reduced to 12% of trials within a session while the percentage of reinforced baseline trials remained at 76%. The remaining 12% of trials within a session were nonreinforced probe trials. The three trial types were pseudo randomly mixed together in a way that prevented more than three consecutive nonreinforced trials.

The nonreinforced probe trials presented color and shape part- and whole-identity problems for which the color and shape of stimuli was novel. The novel colors for the sample and comparison stimuli were pink (Pi), orange (O), and purple (Pu); thus, six color part- and whole-identity problems were created: O → O, not Pi or Pu; Pi → Pi, not O or Pu; and Pu → Pu, not O or Pi. The novel shapes for the sample and comparisons were rhombus' (Rhom), hexagons (Hex), and trapezoids (Trap); thus, six novel shape part- and whole-identity problems were created: Hex → Hex, not Rhom or Trap; Rhom → Rhom, not Hex or Trap; Trap → Trap, not Hex or Rhom. The area of the rhombus', hexagons, and trapezoids was fixed to approximately the same size: 60,028 pixels, 60,840 pixels, and 62,137 pixels, respectively. Figure 3.1.5 shows examples of the novel color and shape part- and whole-identity problems.

To finish Experiment 3, subjects completed six sessions such that 72 nonreinforced probe trials (18 novel color and 18 novel shape part- and whole-identity problems), 72 nonreinforced baseline trials, and 456 reinforced baseline trials were presented to subjects.

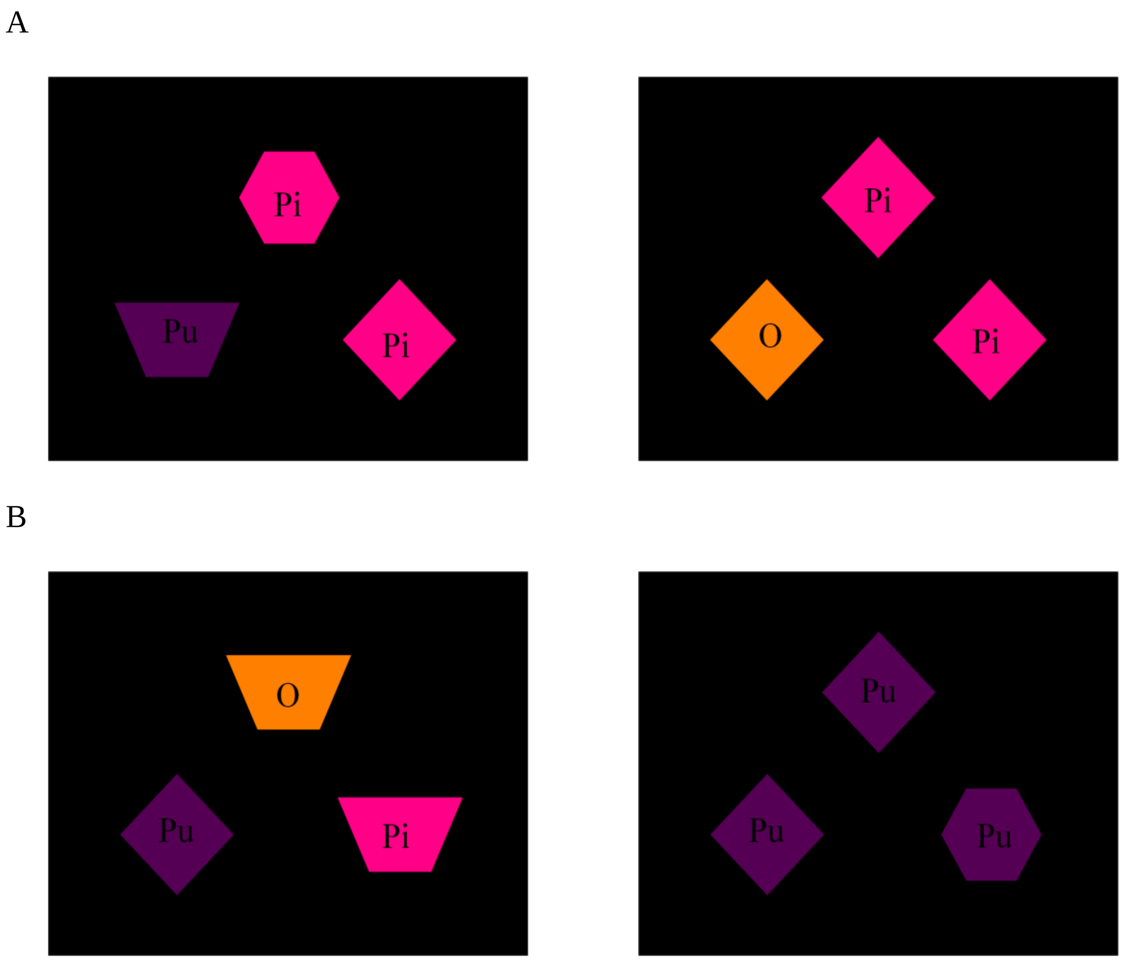


Figure 3.1.5. Example novel color and shape part- and whole-identity problems for the transfer test (Phase 2) of Experiment 3. (A): In the left panel, a color part-identity problem (Pi → Pi, not Pu) and in the right panel, a color whole-identity problem (Pi → Pi, not O). (B) In the left panel, a shape part-identity problem (Trap → Trap, not Rhom) and in the right panel, a shape whole-identity problem (Rhom → Rhom, not Hex).

3.2 Results

3.2.1 Data Analysis

For three experiments subdivided into one to four parts (A, B, C, and D) and/or one to four phases the variables of interest were: (1) subjects (Junior and Madu), (2) problem type (identity and nonidentity), (3) comparison pair type (NOTA absent, as the incorrect comparison, and as the correct comparison), (4) identity problem type (color and shape part- and whole identity), and (5) trial type (nonreinforced baseline, reinforced baseline, and nonreinforced probe trials).

One-tailed binomial tests were used to assess whether the proportion of correct responses for an individual differed from what chance would predict as a function of the aforementioned variables of interest. Chance was set to one-third for three-choice and one-half for two-choice MTS tasks for a number of trials and an alpha level of .05 was used for hypothesis testing of the binomial probability of obtaining k or more correct responses out of n trials.

Chi-square tests of independence (two-tailed) and Fisher's exact tests (one-tailed) were applied to assess the relationship between the proportion of responses (correct vs. incorrect) and the variables of interest ($\alpha = .05$). Cramer's statistic (V) was used to index the strength of the relationship for chi-square tests and the phi coefficient (ϕ) was used to index the strength of the relationship for Fisher's exact tests. One-tailed z tests were used as a follow-up procedure to statistically significant chi-square tests. To maintain the familywise error rate at .05 for each set of pairwise comparisons, a Bonferroni procedure was used to determine the alpha level for each pairwise comparison. Only instances when the adjusted alpha level resulted in a nonsignificant pairwise comparison that would have otherwise been significant are reported indicated in the text.

With respect to response times (reported in seconds), one-tailed correlated groups *t* tests were applied as a function of the aforementioned variables of interest ($\alpha = .05$) with the strength of the relationship indexed by η^2. Casewise deletion was performed when groups had unequal sample sizes. The selection of cases to be deleted was randomly determined so samples remained representative. Response times falling outside of two standard deviations from the mean were deemed outliers and removed from the data before *t* tests were conducted.

The instances when a session did not contain 100 trials are reported in text, tables, or figures. Because multiple sessions were given on a single day, but accuracy (the percentage or proportion of correct responses) not tabulated until all daily sessions were completed, subjects sometimes received additional sessions after reaching the performance criteria; these instances are reported in the text. Regardless of the aforementioned, if the performance criterion was meet then the last 300 trials of a phase or experiment part were considered criterion learning. Additionally, the first 100 trials of a phase or experiment part were considered early learning.

3.2.1.1 Experiment 1A: Color Whole-Identity Responding

3.2.1.1.1 Three-choice MTS (Phase 1)

Madu met the performance criteria (80% correct or more for three consecutive sessions with a minimum of five sessions completed) after completing 500 trials. Madu's accuracy was above chance for the last four sessions (83% correct, $n = 400$; binomial tests, ps $< .001$), but did not differ from chance for the first session (30% correct, $n =$

100; binomial test, p = .272). Figure 3.2.1 displays accuracy as a function of sessions. After completing Phase 1, Madu advanced to Experiment 1B.

Junior did not meet the performance criteria after completing approximately 1,500 trials. His accuracy was 36% correct across all fifteen sessions during which it was above chance for four sessions (49% correct, n = 400; binomial tests, ps < .045), below chance for two sessions (25% correct, n = 200; binomial tests, ps < .046), and not different from chance for nine sessions (33% correct, n = 897; binomial tests, ps > .068). Figure 3.2.1 displays accuracy as a function of sessions. After completing Phase 1, Junior advanced to Phase 2.

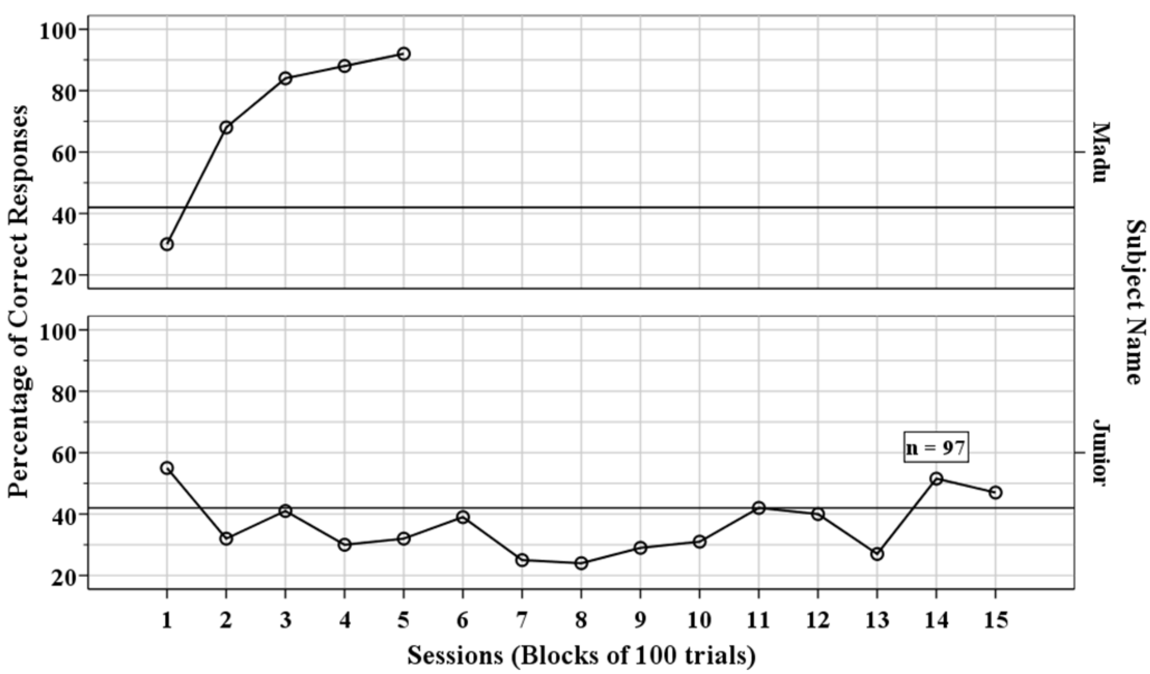


Figure 3.2.1. Subject accuracy as a function of sessions during three-choice color whole-identity MTS tasks in Phase 1. The black horizontal line at 42% correct depicts the lowest percentage of correct responses that was statistically above chance (binomial test, p < .05, n = 100).

3.2.1.1.2 Three-choice MTS with Correction Procedure (Phase 2)

Junior did not meet the performance criteria after completing 1,290 trials. The subject's accuracy was 35% correct for noncorrection trials (n = 428), responses to the initial presentation of a trial, and 31% correct for correction trials (n = 862), responses to representations of a trial following an incorrect response. His accuracy was statistically below chance for one session of noncorrection trials, but otherwise did not differ from chance. Binomial tests and accuracy for the nine sessions are listed in Table 3.2.1. After completing Phase 2, the subject advanced to Phase 3.

Table 3.2.1

Accuracy and Binomial Tests as a function of Sessions for Junior during Three-choice Color Whole-identity MTS tasks under the Correction Procedure in Phase 2

Noncorrection Trials				Correction Trials			
Session	% Correct	*n*	*p*	Session	% Correct	*n*	*p*
1	35.9	39	0.429	1	22.2*	108	0.008
2	36.5	52	0.364	2	31.1	106	0.351
3	38.9	54	0.236	3	36.8	87	0.287
4	37.5	24	0.408	4	29.8	47	0.361
5	33.9	56	0.517	5	31.3	115	0.365
6	34.2	38	0.513	6	29.3	82	0.252
7	33.3	27	0.586	7	32.7	52	0.528
8	37.2	78	0.276	8	33.3	144	0.528
9	26.7	60	0.166	9	36.4	121	0.274

*Below chance accuracy, p < .05.

3.2.1.1.3 Two-choice Three-position MTS (Phase 3)

Junior met the performance criteria after completing approximately 2,000 trials.[10] The subject's accuracy was above chance for the last eight consecutive sessions (79% correct, *n* = 770; binomial tests, *p*s < .035); otherwise, before reaching the performance criterion his accuracy was above chance for six sessions (74% correct, *n* = 592; binomial tests, *p*s < .001) and not different from chance for six sessions (51% correct, *n* = 600; binomial tests, *p*s > .133). The left panel of Figure 3.2.2 displays accuracy as a function of sessions. After completing Phase 3, the subject advanced to Phase 4.

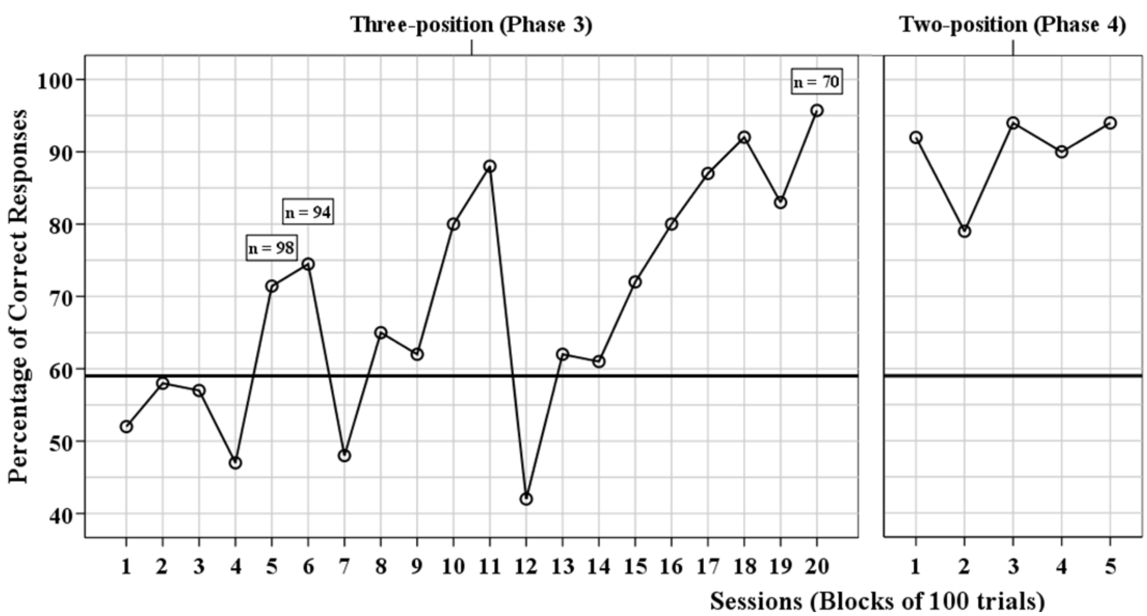


Figure 3.2.2. Accuracy as a function of sessions for Junior during two-choice color whole-identity MTS tasks in Phase 3 and Phase 4. The black horizontal line at 59% correct depicts the lowest percentage of correct responses that was statistically above chance (binomial test, *p* < .05, *n* = 100 and 98).

[10] The subject completed an additional 100-trial session after reaching the performance criteria.

3.2.1.1.4 Two-choice MTS (Phase 4)

Junior met the performance criteria after completing 500 trials. The subject's accuracy averaged 90% correct across all sessions and was above chance for all five sessions (binomial tests, ps < .001). The right panel of Figure 3.2.2 displays accuracy as a function of sessions. After completing Phase 4, the subject advanced to Experiment 1B, Phase 2.

<u>3.2.1.2 Experiment 1B: Concurrent Color Whole-Identity and -Nonidentity Responding</u>

3.2.1.2.1 Two-choice Three-position Modified MTS (Phase 1)

Madu met the performance criteria (i.e., 80% correct or more for three consecutive sessions for both identity and nonidentity problems with a minimum of five sessions completed) after completing 500 trials. For identity problems, the subject's accuracy was above chance for all five sessions (98% correct, n = 250; binomial tests, ps < .001). For nonidentity problems, her accuracy was above chance for the last four consecutive sessions (92% correct, n = 200; binomial tests, ps < .001) and not different from chance for the first session (40% correct, n = 50; binomial test, p = .101). Figure 3.2.3 displays her accuracy as a function of sessions and problem type.

The proportion of correct responses during the first session was examined to evaluate early learning. There was a statistically significant and strong relation between responses and comparison pair type during the first 100 trials, χ^2 (2, N = 100) = 42.86, p < .001, V = .67. Madu's accuracy was lower when NOTA was the correct comparison stimulus (40% correct, n = 50) than when NOTA was the incorrect comparison (100% correct, n = 24; z = 4.92, p < .001) and when NOTA was absent as a comparison (100%

correct, $n = 26$; $z = 5.08$, $p < .001$); also, accuracy did not differ between when NOTA was absent as a comparison and when NOTA was incorrect.

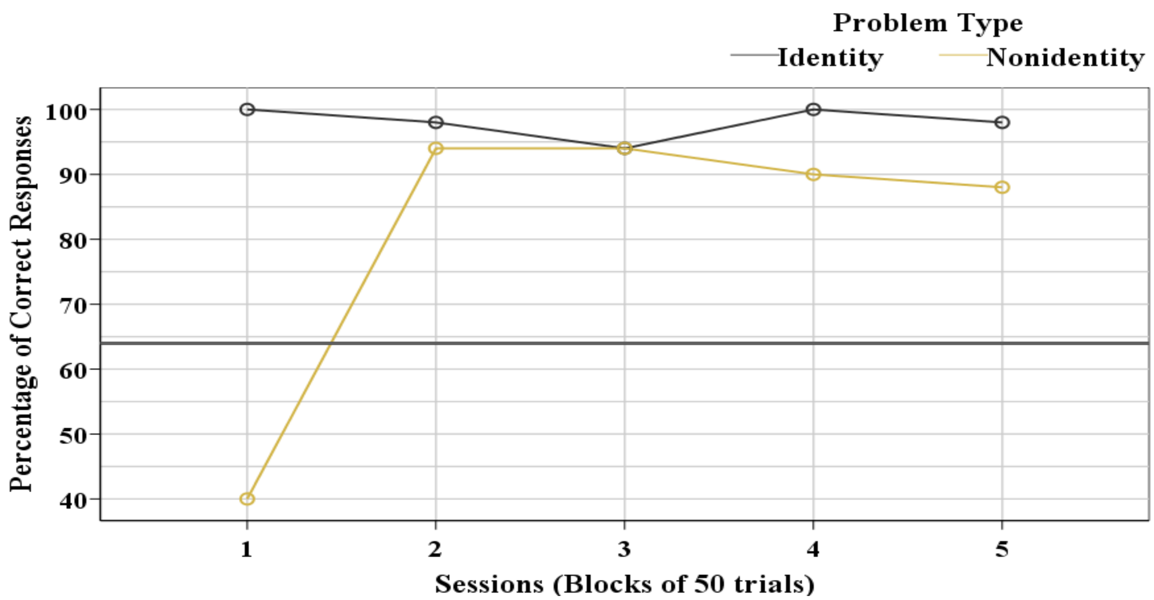


Figure 3.2.3. Accuracy as a function of sessions and problem type for Madu during two-choice three-position concurrent color whole-identity and -nonidentity modified MTS tasks in Phase 1. The black horizontal line at 64% correct depicts the lowest percentage of correct responses that was statistically above chance (binomial test, $p < .05$, $n = 50$).

3.2.1.2.2 Two-choice Modified MTS (Phase 2)

Madu met the performance criteria after completing 500 trials. Her accuracy was statistically above chance for both identity (98% correct) and nonidentity problems (91% correct) during each of the five sessions. Table 3.2.2 displays the subject's accuracy and results of binomial tests.

Table 3.2.2

Accuracy and Binomial Tests for Two-choice Concurrent Color Whole-Identity and -Nonidentity Responding in Phase 2 for Madu

Identity Problems				Nonidentity Problems			
Session	% Correct	*N*	*P*	Session	% Correct	*n*	*p*
1	98.3*	58	0.001	1	97.6*	42	0.001
2	100.0*	52	0.001	2	81.3*	48	0.001
3	98.7*	75	0.001	3	100.0*	25	0.001
4	93.3*	75	0.001	4	80.0*	25	0.002
5	98.7*	75	0.001	5	96.0*	25	0.001

*Above chance accuracy, $p < .05$.

With his first experience with the two-choice modified MTS task, Junior met the performance criteria after completing 800 trials. For identity problems, his accuracy was above chance for the last five consecutive sessions (90% correct, $n = 250$; binomial tests, *p*s < .001); otherwise, his accuracy was not different from chance for two sessions (46% correct, $n = 104$; binomial tests, *p*s = .339) and above chance for one other session (94% correct, $n = 50$; binomial test, $p < .001$). For nonidentity problems, Junior's accuracy was above chance for the last three consecutive sessions (90% correct, $n = 150$ trials; binomial tests, *p*s < .001); otherwise, his accuracy was below chance for three sessions (10% correct, $n = 150$; binomial tests, *p*s < .001) and not different from chance for two sessions (53% correct, $n = 96$; binomial tests, $ps > .333$). Figure 3.2.4 displays his accuracy as a function of sessions and problem type.

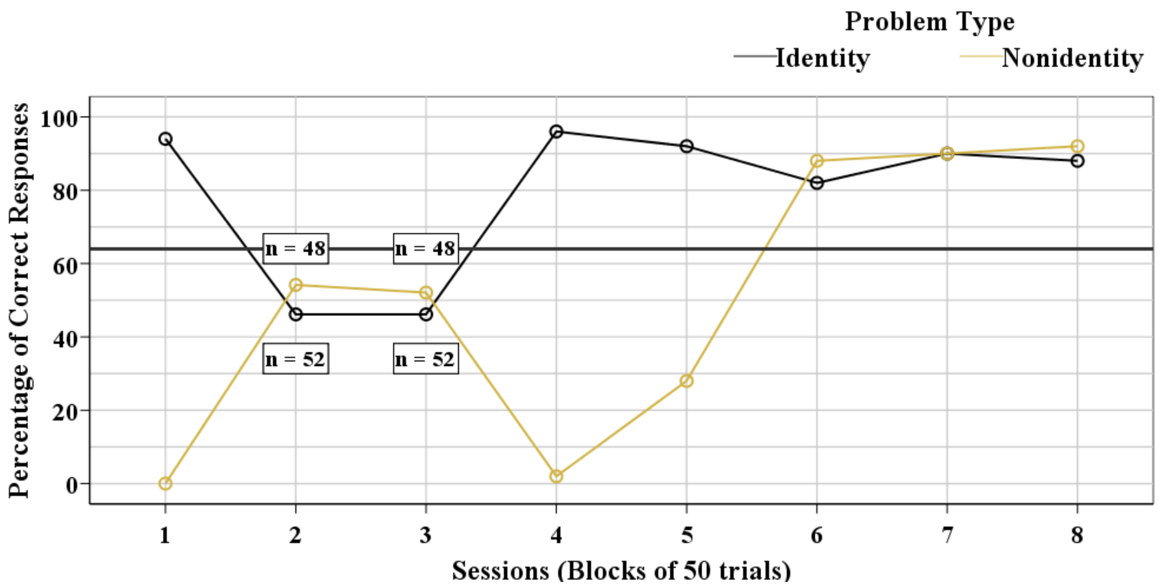


Figure 3.2.4. Accuracy as a function of sessions and problem type for Junior during two-choice concurrent color whole-identity and -nonidentity modified MTS tasks in Phase 2. The black horizontal line at 64% correct shows the lowest percentage of correct responses that was statistically above chance (binomial test, $p < .05$, $n = 50$).

For Junior, the proportion of correct responses during the first session was examined to evaluate early learning.[11] There was a statistically significant and strong relation between responses and comparison pair type during the first 100 trials, χ^2 (2, N = 100) = 89.35, $p < .001$, $V = .95$. Junior's accuracy was lower when NOTA was the correct comparison stimulus (0% correct, $n = 50$) than when NOTA was the incorrect comparison (100% correct, $n = 24$; $z = 8.60$, $p < .001$) and when NOTA was absent as a

[11] Early learning was not evaluated with Madu as her first exposure to concurrent color whole-identity and -nonidentity responding was in the previous phase.

comparison (88% correct, n = 26; z = 7.96, p < .001); also, accuracy did not differ between when NOTA was absent as a comparison and when NOTA was the correct comparison (z = 1.72, p = .086).

The proportion of correct responses during the final three sessions was examined to evaluate criterion learning for both Madu and Junior. There was a statistically significant and moderate relation between responses and comparison pair type during the last 300 trials for both Madu, χ^2 (2, N = 300) = 9.79, p = .008, V = .18, and Junior, χ^2 (2, N = 300) = 10.12, p = .006, V = .18. Madu's accuracy was higher when NOTA was absent as a comparison stimulus (99% correct, n = 150) than when NOTA was the correct comparison (92% correct, n = 75; z = 2.99, p < .001) and when NOTA was the incorrect comparison (92% correct, n = 75; z = -2.99, p < .001); also, accuracy did not differ between when NOTA was the incorrect and correct comparison stimulus. Junior's accuracy was higher when NOTA was absent as a comparison stimulus (95% correct, n = 75; z = -2.88, p = .002) and when NOTA was the correct comparison (90% correct, n = 150; z = -2.33, p = .010) than when NOTA was the incorrect comparison (79% correct, n = 75); also, accuracy did not differ between when NOTA was absent as a comparison and when NOTA was the correct comparison (z = 1.19, p = .118).

Finally, responding speed during the final three sessions was examined for both Madu and Junior. Madu responded more slowly when NOTA was the correct comparison (M = 1.20 s, SE = .03, n = 69) than when NOTA was the incorrect comparison (M = .83 s, SE = .03, n = 69 after randomly deleting 5 cases; t[68] = -8.70, p < .001, η^2 = .53) and when NOTA was absent as a comparison identity (M = .85 s, SE = .02, n = 69 after randomly deleting 73 cases; t[68] = -8.75, p < .001, η^2 = .53), but there was no difference in response speed between when NOTA was the incorrect comparison and when NOTA was absent as a comparison (t[68] = -.66, p = .513). Likewise, Junior responded more slowly when NOTA was the correct comparison (M = 1.98 s, SE = .08, n = 75 after randomly deleting 74 cases) than when NOTA was the incorrect comparison (M = 1.64 s, SE = .06, n = 75; t[74] = -3.45, p = .001, η^2 = .14) and when NOTA was absent as a comparison (M = 1.64 s, SE = .06, n = 75; t[74] = -3.60, p = .001, η^2 = .15), but there was no difference in response speed between when NOTA was the incorrect comparison and when NOTA was absent as a comparison (t[74] = -.02, p = .985).

3.2.1.3 Experiment 2A: Concurrent Color and Shape Part-Identity Responding

Both subjects failed to meet the performance criteria after completing 3,000 trials because their accuracy was not above 80% correct for three consecutive sessions for both color and shape part-identity problems.

For color part-identity problems, Madu's responses were 91% correct across all sessions (n = 1,500) and her accuracy was above chance for each of the 30 sessions (binomial tests, ps < .016). Similarly, Junior's responses were 81% correct across all sessions (n = 1,500) and his accuracy was above chance for 28 sessions (binomial tests, ps < .032) and no different from chance for 2 sessions (binomial tests, ps > .059).

On the other hand, for shape part-identity problems, Madu's responses were 67% correct across all sessions (n = 1,500 trials) with her accuracy above chance for 16 sessions (74% correct, n = 800; binomial tests, ps < .032) and at chance for 14 sessions (59% correct, n = 700; binomial tests, ps > .059). Similarly, Junior's responses were 57% correct across all sessions (n = 1,500) with his accuracy above chance for eight sessions (74% correct, n = 400; binomial tests, ps < .032) and at chance for 22 sessions (52%

correct, n =1,100; binomial tests, ps > .059). Figure 3.2.5 displays subject accuracy as a function of sessions and identity problem type.

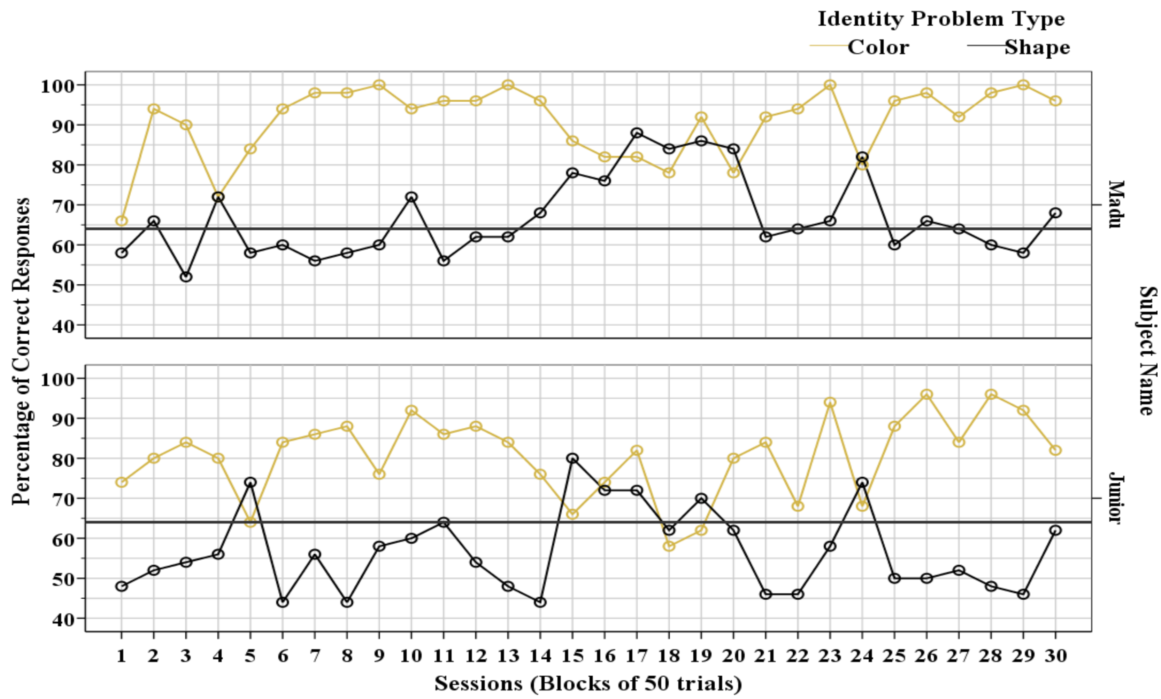


Figure 3.2.5. Subject accuracy as a function of sessions and identity problem type during concurrent color and shape part-identity MTS tasks in Experiment 2A. The black horizontal line at 64% correct shows the lowest percentage of correct responses that was statistically above chance (binomial test, $p < .05$, $n = 50$).

3.2.1.4. Experiment 2B: Shape Part- and Whole-Identity Responding

3.2.1.4.1. Shape Part-Identity (Phase 1)

Madu met the performance criteria after completing 500 trials.[12] Her accuracy was above chance for all five sessions (77% correct, n = 500; binomial tests, ps < .003). The left panel of Figure 3.2.6 displays accuracy as a function of each of the five sessions. After completing Phase 1, Madu advanced to Experiment 2C.

On the other hand, Junior failed to meet the performance criteria after completing 1,500 trials. His accuracy was above chance during one session (62% correct, n = 100; binomial, p = .010); otherwise, his accuracy was not different from chance (53% correct, n = 1,400; binomial tests, ps > .067). The left panel of Figure 3.2.6 displays accuracy as a function of each session. After completing Phase 1, Junior advanced to Phase 2.

3.2.1.4.2. Shape Whole-Identity (Phase 2)

Junior met the performance criteria after completing 2,000 trials. His accuracy was above chance for the last 14 consecutive sessions (71% correct, n = 1,400; binomial tests, ps < .044); otherwise, his accuracy was at chance for six sessions (56% correct, n = 600; binomial tests, ps > .067). The second to left panel of Figure 3.2.6 displays accuracy as a function of sessions. After completing Phase 2, Junior advanced to Phase 3.

3.2.1.4.3. Shape Part-Identity (Phase 3)

Junior met the performance criteria after completing 1,200 trials. His accuracy was above chance for all 12 sessions (76% correct; binomial tests, ps < .001). The second to right panel of Figure 3.2.6 displays accuracy as a function of sessions. After completing Phase 3, Junior advanced to Experiment 2C.

3.2.1.4.4. Early and Criterion Learning Between Experiment Phases

---

[12] Ibid footnote 10, pg. 78.

Comparing the proportion of correct responses during the last 300 trials of shape part-identity responding in Phase 1 for Madu to the last 300 trials of shape part-identity learning in Phase 3 for Junior revealed that criterion learning accuracy did not differ between Madu and Junior (both 84% correct, *n*s = 300; Fisher's exact test, $p = .50$).

Additionally, for Junior, the proportion of correct responses during criterion learning of shape whole-identity in Phase 2 (84% correct, $n = 300$) was statistically higher than the proportion of correct responses during early learning of shape part-identity responding in Phase 3 (71% correct, $n = 100$), Fisher's exact test, $p = .005$, $\phi = .138$.

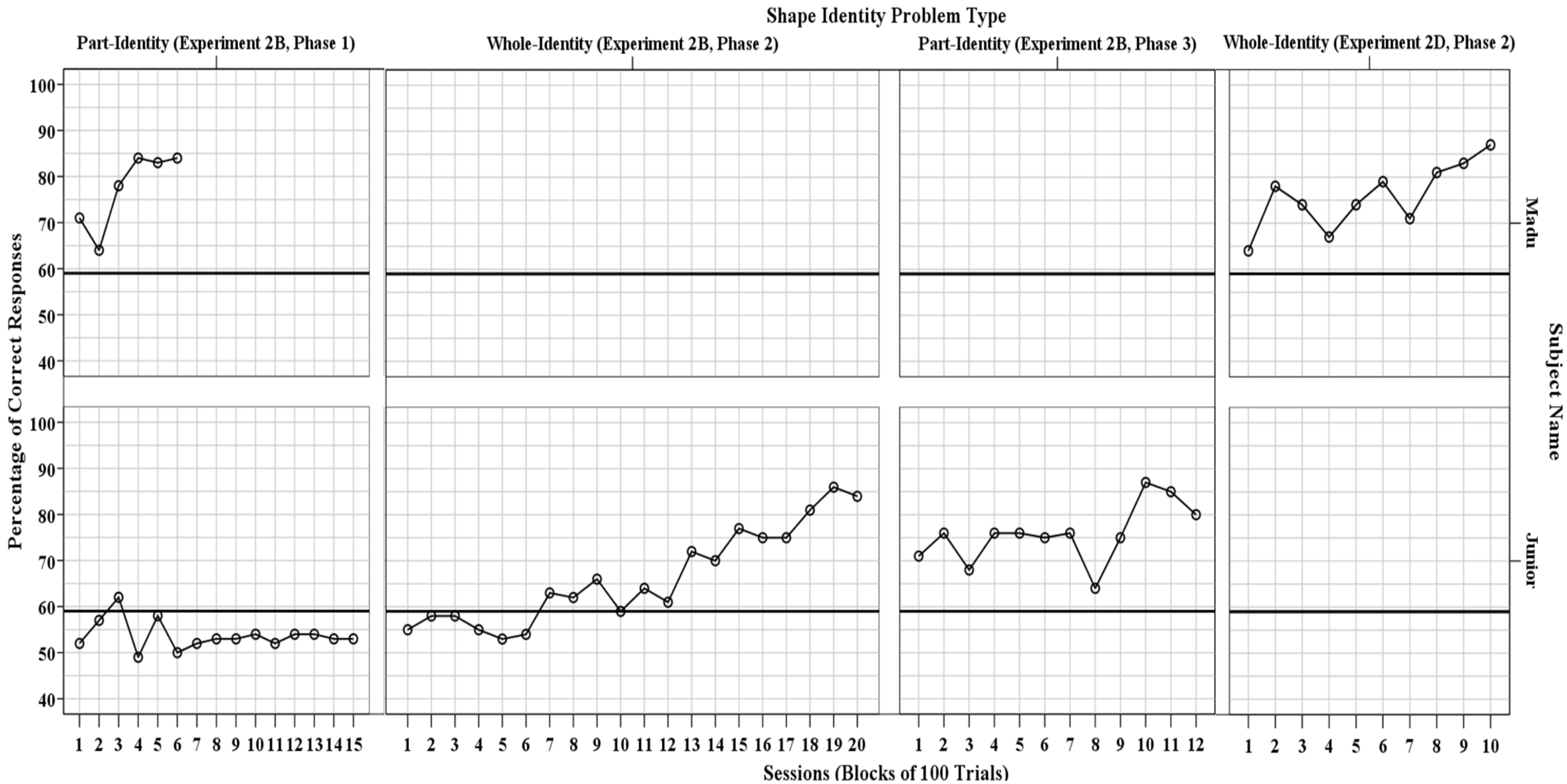


Figure 3.2.6. Subject accuracy as a function of sessions for shape part- and whole-identity MTS tasks in Experiment 2B and 2D. The black horizontal line at 59% correct shows the lowest percentage of correct responses that was statistically above chance (binomial test, $p < .05$, $n = 100$). For Madu, note that color and shape part-identity responding in Experiment 2C and color whole-identity responding in Experiment 2D occurred before shape whole-identity responding.

3.2.1.5. Experiment 2C: Concurrent Color and Shape Part-Identity Responding

Madu met the performance criteria after completing 1,200 trials.[13] For color part-identity problems, her accuracy was above chance for all 13 sessions (97% correct, n = 650; binomial tests, ps < .001). Likewise, for shape part-identity problems, her accuracy was above chance for all 13 thirteen sessions (75% correct, n = 650; binomial tests, ps < .032). Junior met the performance criteria after completing 1,500 trials. For color part-identity problems, his accuracy was above chance for all fifteen sessions (86% correct, n = 650; binomial tests, ps < .001). Likewise, for shape part-identity problems, his accuracy was above chance for all 15 sessions (79% correct, n = 650; binomial tests, ps < .003). Figure 3.2.7 displays subject accuracy as a function of sessions and identity problem type.

Performance during the last three sessions was examined to evaluate criterion learning. Madu was more accurate with color (98% correct, n = 150) than shape (79% correct, n = 150) part-identity problems, Fisher's exact test, $p < .001$, $\phi = .301$). On the other hand, Junior' accuracy with color (83% correct, n = 150) and shape (88% correct, n = 150) part-identity problems was not different, Fisher's exact test, $p = .161$. Further, Madu's accuracy was significantly higher than Junior's for color part-identity problems (Fisher's exact test, $p < .001$, $\phi = .252$) and Junior's was significantly higher than Madu's for shape part-identity problems (Fisher's exact test, $p < .022$, $\phi = .125$).

Finally, Madu responded faster to color (M = .80 s, SE = .01, n = 146 with 2 cases deleted) than shape (M = 1.12 s, SE = .03, n = 146) part-identity problems during criterion learning, $t(145) = -11.76$, $p < .001$, $\eta^2 = .49$. On the other hand, Junior

[13] Ibid footnote 10, pg. 78.

responded no differently to color (M = 1.36 s, SE = .044, n = 143) and shape (M = 1.31 s, SE = .04, n = 143) part-identity problems during criterion learning, $t(142) = .779$, p = .219.

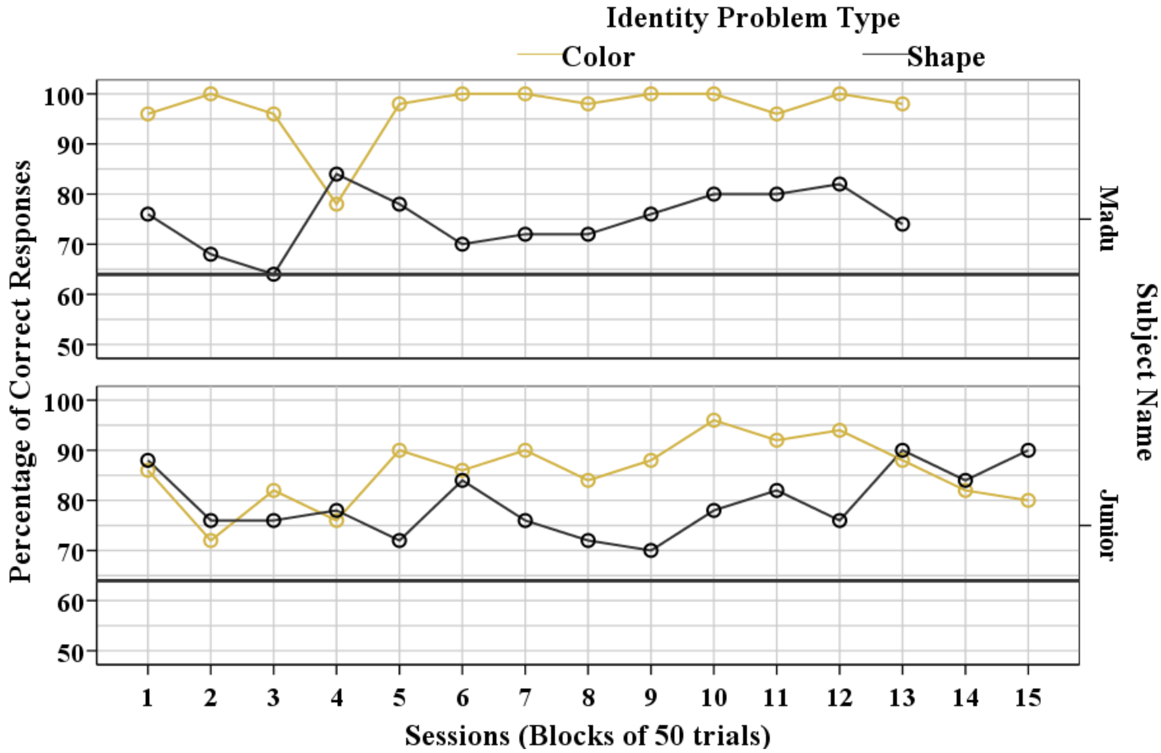


Figure 3.2.7. Subject accuracy as a function of sessions and identity problem type during concurrent color and shape part-identity MTS tasks. The black horizontal line at 64% correct shows the lowest percentage of correct responses that was statistically above chance (binomial test, $p < .05$, n = 50).

3.2.1.5.1. Early and Criterion Learning Between Experiment Parts

In relation to shape part-identity problems, Madu's accuracy was not different during criterion learning in Experiment 2B, Phase 1 when only shape part-identity problems were presented (84% correct, n = 300) and early learning in Experiment 2C when both shape and color part-identity problems were presented (76% correct, $n = 150$), Fisher's exact test, $p =.132$. Likewise, Junior's accuracy for shape part-identity problems was not different during criterion learning in Experiment 2B, Phase 3 when only shape part-identity problems were presented (84% correct, $n = 300$) and early learning in Experiment 2C when both shape and color part-identity problems were presented (88% correct, $n = 150$), Fisher's exact test, $p = .314$.

Furthermore, the same pattern was evident for both subjects in relation to color part-identity problems. That is, neither Madu's nor Junior's criterion accuracy for color part-identity problems in Experiment 2A differed from their early learning in Experiment 2C (Madu: 98% vs. 96% correct, respectively; Junior: 90% vs. 86% correct, respectively), Fisher's exact tests, *p*s > .293.

<u>3.2.1.6. Experiment 2D: Color and Shape Whole-Identity Responding</u>

3.2.1.6.1. Color Whole-Identity (Phase 1)

Madu met the performance criteria (i.e., at least 80% correct for three consecutive sessions) after completing 300 trials. Her accuracy was above chance for all three sessions (all 100% correct, *n*s = 100; binomial tests, *p*s < .001). After completing Phase 1, she advanced to Phase 2. Similarly, Junior met the performance criteria after completing 300 trials.[14] His accuracy was above chance for all four sessions (in order by session: 95%, 92%, 92%, and 95% correct, *n*s = 95, 92, 92, and 95; binomial tests, *p*s < .001). After completing Phase 1, he advanced to Experiment 3.

3.2.1.6.2. Shape Whole-Identity (Phase 2)

---

[14] Ibid footnote 10, pg. 78.

Madu met the performance criteria after completing 1,000 trials. Her accuracy was above chance for all ten sessions (76% correct; n = 1,000; binomial tests, ps < .003). The right panel of Figure 3.2.6 displays Madu's accuracy as a function of sessions.

3.2.1.7. Experiment 3: Concurrent Color and Shape Part- and Whole-Identity Responding and Test of Transfer

3.2.1.7.1. Baseline (Phase 1)

Madu met the performance criteria (i.e., at least 70% correct for color part-identity, color whole-identity, shape part-identity, and shape whole-identity problems for three consecutive sessions with a minimum of five sessions completed) after completing 1,000 trials.[15] Her accuracy was above chance during all 12 sessions of color part- and whole-identity problems (98% and 99% correct, respectively, ns = 300; binomial tests, ps < .001). For shape part-identity problems, her accuracy was above chance for the final five consecutive sessions (78% correct, n = 125; binomial tests, ps < .022); otherwise, her accuracy was above chance for two other sessions (72% correct, n = 50; binomial tests, ps = .022) and at chance for five sessions (62% correct, n = 125; binomial tests, ps > .054). For shape whole-identity problems, her accuracy was above chance for the final seven consecutive sessions (79% correct, n = 175; binomial tests, ps < .007); otherwise, her accuracy was above chance for three sessions (83% correct, n = 75; binomial tests, ps < .022) and no different from chance for two sessions (64% correct, n = 50; binomial tests, ps > .054). The left panel of Figure 3.2.8 displays Madu's accuracy as a function of sessions and trial type. After completing Phase 1, she advanced to Phase 3.

---

[15] Madu received two additional 100-trial sessions after she reached the criteria.

Junior met the performance criteria after completing 1,700 trials. His accuracy was above chance for all 17 sessions for color part- and whole-identity problems (93% and 95% correct, respectively, *n*s = 425; binomial tests, *p*s < .002). For shape part-identity, his accuracy was above chance for the final four consecutive sessions (80% correct, *n* = 100; binomial tests, *p*s < .007); otherwise, his accuracy was above chance for four other sessions (75% correct, *n* = 100; binomial tests, *p*s < .007) and at chance for nine sessions (65% correct, *n* = 225; binomial tests, *p*s > .054). For shape whole-identity problems, his accuracy was above chance for the final three consecutive sessions (84% correct, *n* = 75; binomial tests, *p*s < .002); otherwise, his accuracy was above chance for seven sessions (81% correct, *n* = 175; binomial tests, *p*s < .022) and no different from chance for seven sessions (62% correct, *n* = 175; binomial tests, *p*s > .054). The left panel of Figure 3.2.9 displays Junior's accuracy as a function of sessions and trial type. After completing Phase 1, he advanced to Phase 2.

The final three sessions of reinforced baseline trials were used to assess criterion learning. For Madu, the relationship between responses and identity problem type was statistically significant and moderately strong, χ^2 (3, *N* = 300) = 28.62, *p* < .001, *V* = .31. The pattern was such that her accuracy did not differ between color part- and whole-identity problems (96% vs. 99% correct, respectively; *z* = 1.01, *p* = .155, *n* = 75) or between shape part- and whole-identity problems (77% vs. 76% correct, respectively; *z* = .19, *p* = .423, *n* = 75), but her accuracy was higher for color part- and whole-identity problems than shape part- and whole-identity problems (*z*s > 3.36, *p*s < .001, *n*s = 75). For Junior, on the other hand, the relationship between responses and identity problem type was not statistically significant, χ^2 (3, *N* = 300) = 6.53, *p* = .088, which indicates that his accuracy did not differ statistically among color part-identity (89% correct), color whole-identity (92% correct), shape part-identity (79% correct), and shape whole-identity (84% correct) problems during the final three sessions (*n*s = 75).

With respect to response latency during the last three sessions of reinforced baseline trials, statistically significant correlated groups *t* tests indicated that both subjects responded faster to color whole-identity problems (Madu: $M = 1.16$ s, $SE = .07$, $n = 75$; Junior: $M = 1.26$ s, $SE = .07$, $n = 150$) than shape part-identity problems (Madu: $M = 1.33$ s, $SE = .05$, $n = 75$; Junior: $M = 1.46$ s, $SE = .41$, $n = 150$), but the effect was weak for both Madu, $t(74) = 1.89$, $p = .031$, $\eta^2 = .05$, and Junior, $t(149) = 2.70$, $p = .004$, $\eta^2 = .05$. No other paired comparison was statistically significant for Madu or Junior: $ts(74) < 1.42$, $ps > .081$ for Madu; $ts(149) < 1.15$, $ps > .126$ for Junior.

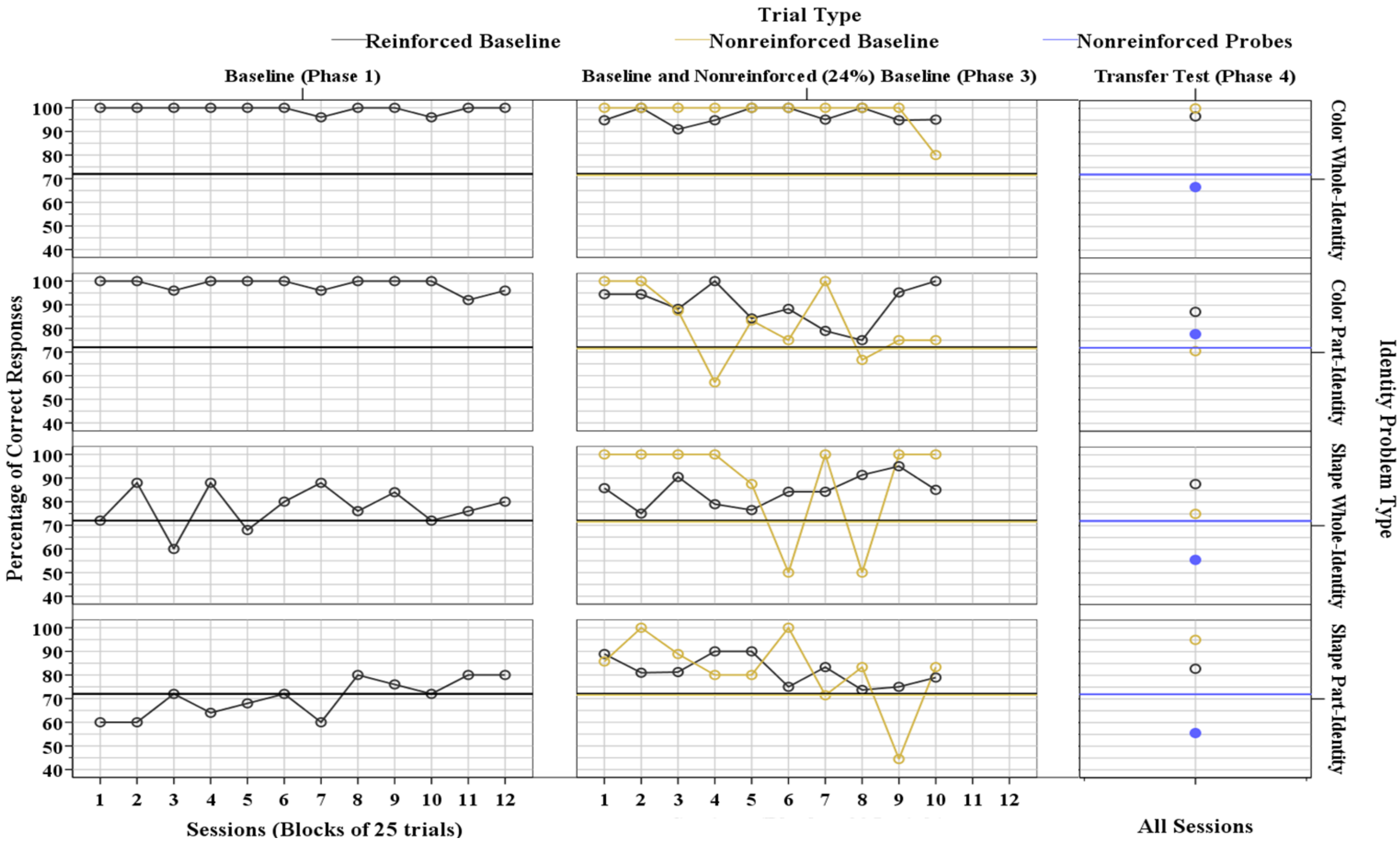


Figure 3.2.8. Accuracy as a function of sessions and trial type for each identity problem type and phase of Experiment 3 for Madu. The black horizontal line at 72% correct shows the lowest percentage of correct responses that was statistically above chance for reinforced baseline trials in Phase 1 (binomial test, $p < .05$, $n = 25$); likewise, with the black and gold line for the average accuracy of reinforced and nonreinforced baseline trials in Phase 2 (binomial test, $p < .05$, $n = 25$) and with the blue line for nonreinforced probe trials in Phase 4 (binomial test, $p < .05$, $n = 18$).

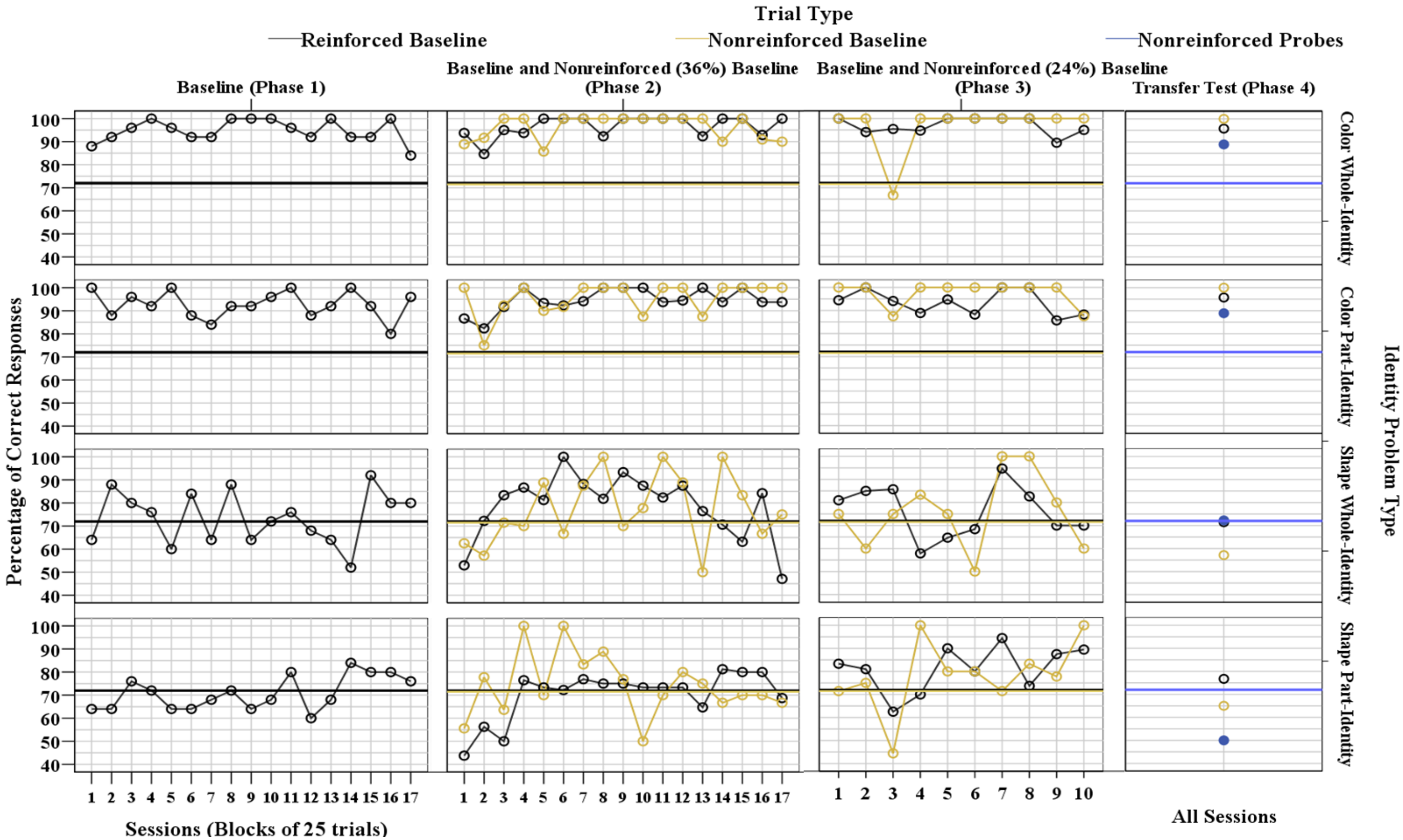


Figure 3.2.9. Accuracy as a function of sessions and trial type for each identity problem type and phase of Experiment 3 for Junior. The black horizontal line at 72% correct shows the lowest percentage of correct responses that was statistically above chance for reinforced baseline trials in Phase 1 (binomial test, $p < .05$, $n = 25$); likewise, with the black and gold line for the average accuracy of reinforced and nonreinforced baseline trials in Phase 2 and 3 (binomial test, $p < .05$, $n = 25$) and with the blue line for nonreinforced probe trials in Phase 4 (binomial test, $p < .05$, $n = 18$).

3.2.1.7.2. Baseline and Nonreinforced (36%) Baseline (Phase 2)

Junior met the performance criteria after completing 600 trials. His accuracy was above chance for all six sessions of color part- and whole-identity problems (91% and 95% correct, respectively, *n*s = 150; binomial tests, *p*s < .05). For shape part-identity problems, his accuracy was above chance for the last three consecutive sessions (79% correct, *n* = 75; binomial tests, *p*s < .002); otherwise, his accuracy was not different from chance for three sessions (56% correct, *n* = 75; binomial tests, *p*s > .115). For shape whole-identity problems, his accuracy was above chance for the last four consecutive sessions (83% correct, *n* = 100; binomial tests, *p*s < .002); otherwise, his accuracy was not different from chance for two sessions (62% correct, *n* = 50; binomial tests, *p*s > .054).

During the 11 additional sessions that he received after reaching the performance criteria, Junior's responses were 97% and 98% correct (*n*s = 275) for color part- and whole-identity problems, respectively, and he was above chance for each of the 11 sessions (binomial tests, *p*s < .001). The subject's responses were 74% and 79% correct (*n*s = 275) for shape part- and whole-identity problems during the 11 additional sessions that he received after reaching the criteria. His accuracy was above chance for eight of the 11 sessions for both shape part- and whole-identity problems (74% and 85% correct, respectively, *n*s = 200; binomial tests, *p*s < .002); otherwise his accuracy was not different for chance for three sessions for shape part- and whole-identity problems (67% and 64% correct, respectively, *n*s = 200; binomial tests, *p*s > .054). The second to left panel of Figure 3.2.9 displays Junior's accuracy as a function of sessions and trial type.

3.2.7.1.3. Baseline and Nonreinforced (24%) Baseline (Phase 3)

Madu met the performance criteria after completing 500 trials. For color part- and whole-identity problems, her accuracy was above chance for all five sessions (90% and 97% correct, respectively, *n*s = 125; binomial tests, *p*s < .001). For shape part- and whole-identity problems, Madu's accuracy was above chance for all five sessions (86% and 85% correct, respectively, *n*s = 125; binomial tests, *p*s < .002). During the five additional sessions that she received after reaching the performance criteria, Madu's accuracy above chance for all five sessions for color part-identity, color whole-identity, and shape whole-identity problems (85%, 97%, 79% correct, respectively, *n*s = 125; binomial tests, *p*s < .022). Her accuracy was above chance during four of the five sessions for shape part-identity problems (79% correct, $n = 100$; binomial tests, *p*s < .008); otherwise, her accuracy did not differ from chance for one session (64% correct, $n = 25$; binomial test, $p = .115$). The middle panel of Figure 3.2.8 displays Madu's accuracy as a function of sessions and trial type. After completing Phase 3, the subject advanced to Phase 4 transfer testing.

Junior met the performance criteria after completing 900 trials. His accuracy was above chance for all ten sessions for color part- and whole-identity problems (94% and 97% correct, respectively, *n*s = 250; binomial tests, *p*s < .001). For shape part-identity problems, his accuracy was above chance for the last seven consecutive sessions (83% correct, $n = 175$; binomial tests, *p*s < .008); otherwise, his accuracy was above chance for two other sessions (80% correct, $n = 50$; binomial tests, *p*s < .002) and not different from chance for one session (56% correct, $n = 25$ trials; binomial test, $p > .345$). For shape whole-identity problems, his accuracy was above chance for six sessions (83 % correct, $n = 150$; binomial tests, *p*s < .022); otherwise, his accuracy was at chance for four sessions

(66% correct, n = 100; binomial tests, ps > .054). The second to right panel of Figure 3.2.9 displays Junior's accuracy as a function of sessions and trial type. After completing Phase 3, the subject advanced to Phase 4 transfer testing.

Accuracy for nonreinforced baseline trials during the last three sessions was used to evaluate criterion learning. Madu's accuracy was above chance for nonreinforced baseline trials of color part-identity, color whole-identity, and shape whole-identity problems (71%, 94%, and 92% correct, ns = 21, 18, and 12, respectively; binomial tests, ps < .039), but not above chance for nonreinforced baseline trials of shape part-identity problems (67% correct, n = 21; binomial test, p = .095). Junior's accuracy was above chance for nonreinforced baseline trials of color part-identity, color-whole identity, and shape part-identity problems (95%, 100%, and 86% correct, respectively, ns = 21, 18, and 21; binomial tests, ps < .001), but not above chance for nonreinforced baseline trials of shape whole-identity problems (75% correct, n = 12; binomial tests, ps = .073).

Further, for nonreinforced baseline trials, the relationship between responses and identity problem type was statistically significant and moderately strong during criterion learning for both Madu, χ^2 (3, N = 300) = 17.94, p < .001, V = .26, and Junior, χ^2 (3, N = 300) = 17.31, p < .001, V = .24. For Madu, the pattern was such that her accuracy was lower for shape part-identity problems (73% correct) than color part-identity (85% correct; z = 1.81, p = .035), shape whole-identity (91% correct; z = 2.76, p = .003), and color whole-identity (96% correct; z = 3.85, p < .001); additionally, her accuracy for color part-identity problems was lower than that for color whole-identity (z = 2.25, p = .012. No other pairwise comparison was significant (zs = 1.31 and 1.01, ps = .095 and .157). For Junior, the pattern was such that his accuracy for shape whole identity problems (75% correct) was significantly lower than his accuracy for color whole-identity (96% correct; $z = 3.69$, $p < .001$) and color part-identity (92% correct; $z = -2.85$, $p = .002$) and his accuracy for shape part-identity problems (84% correct) was significantly lower than that for color whole-identity ($z = 2.45$, $p = .007$). No other pairwise comparison was significant (zs = -1.41, 1.03, and 1.51, ps > .066).

3.2.1.7.4. Transfer Test (Phase 4)

The right panel of Figure 3.2.8 displays Madu's accuracy during the transfer of learning test. Performance was evaluated across all six transfer test sessions for reinforced baseline trials. Madu's accuracy was above chance for color part- and whole-identity problems (87% and 97% correct, respectively, ns = 117 and 118; binomial tests, ps < .001) and for shape part- and whole-identity problems (83% and 88% correct, respectively, ns = 116 and 105; binomial tests, ps < .001).

Performance was also evaluated across all six transfer test sessions for nonreinforced baseline trials. Madu's accuracy for nonreinforced baseline trials was above chance for color whole-identity problems (100% correct, $n = 7$; binomial test, $p = .008$) and shape part- and whole-identity problems (95% and 75% correct, respectively, ns = 20 and 28; binomial tests, ps < .007), but did not differ from chance for color part-identity problems (71% correct, $n = 17$; binomial test, $p = .072$).

Finally, performance was evaluated across all six transfer test sessions for nonreinforced probe trials and then compared to criterion learning in Phase 3. First, Madu's accuracy was above chance for novel color part-identity problems (78% correct, $n = 18$; binomial test, $p = .016$). Additionally, the relationship between responses to color part-identity problems and nonreinforced probe trials (78% correct), nonreinforced baseline trials during criterion learning in Phase 3 (71% correct), and reinforced baseline trials during criterion learning in Phase 3 (91% correct) was not statistically significant, χ^2 (2, N = 93) = 4.78, p = .092, which revealed that her accuracy with novel color part-identity problems was not different from her accuracy with familiar color part-identity problems.

Second, Madu's accuracy did not differ from chance for nonreinforced probe trials of the novel shape part-identity problems (56% correct, n = 18; binomial test, p = .407). Additionally, the relation between responses to shape part-identity problems and nonreinforced probe trials, nonreinforced baseline trials during criterion learning in Phase 3 (67% correct), and reinforced baseline trials during criterion learning in Phase 3 (76% correct) was not statistically significant, χ^2 (2, N = 93) = 2.80, p = .247, which revealed that response accuracy for novel shape part-identity problems was not different from criterion accuracy for familiar shape part-identity problems.

Last, Madu's accuracy did not differ from chance for nonreinforced probe trials of the novel color whole-identity problems (67% correct, n = 18; binomial test, p = .119) and the novel shape whole-identity problems (56% correct, n = 18; binomial test, p = .407). In addition, the relation between responses to color whole-identity problems and responses to shape whole-identity problems and the nonreinforced probe trials, nonreinforced baseline trials during criterion learning in Phase 3, and reinforced baseline trials during criterion learning in Phase 3 was statistically significant: χ^2s (2, Ns = 93) > 13.24, ps < .001, Vs > .38. For color whole-identity problems, the pattern was such that Madu's accuracy was significantly lower for nonreinforced probe trials (67% correct) than nonreinforced baseline trials (94% correct, n = 18; z = 3.57, p < .001) and reinforced baseline trials (96% correct, $n = 57$; $z = 2.11$, $p = .018$), and accuracy did not differ between nonreinforced and reinforced baseline trials ($z = 3.86$, $p = .350$), which revealed that response accuracy for novel color whole-identity problems was lower than criterion accuracy for familiar color whole-identity problems. For shape whole-identity problems, Madu's accuracy was significantly lower for nonreinforced probe trials (56% correct) than nonreinforced baseline trials (92% correct, $n = 11$; $z = 2.11$, $p < .017$) and reinforced baseline trials (90% correct; $n = 57$, $z = 3.46$, $p < .001$), and accuracy did not differ between nonreinforced and reinforced baseline trials ($z = -1.30$, $p = .448$), which reveals that response accuracy for novel shape whole-identity problems was lower than criterion response accuracy for familiar shape whole-identity problems.

The right panel of Figure 3.2.9 displays Junior's accuracy during the transfer test. Performance was evaluated across all six transfer test sessions for nonreinforced baseline trials. Junior's accuracy was above chance for color part- and whole-identity problems (both 100% correct; *n*s = 17 and 7, respectively; binomial tests, *p*s < .008), but Junior's accuracy was not different from chance for shape part- and whole-identity problems (65% and 57% correct, *n*s = 20 and 28, respectively; binomial tests, *p*s > .135). Performance was also evaluated across all six transfer test sessions for reinforced baseline trials. Junior's accuracy was above chance for reinforced baseline trials of color part- and whole-identity problems (96% and 96% correct, *ns* = 117 and 118, respectively; binomial tests, *p*s < .001) and reinforced baseline trials of shape part- and whole-identity problems (77% and 71% correct, *n*s = 116 and 105, respectively; binomial tests, *p*s < .001).

Finally, performance was evaluated across all six transfer test sessions for nonreinforced probe trials and then compared to criterion learning in Phase 3. For nonreinforced probe trials of the novel color part-identity, color whole-identity, and shape whole-identity problems, Junior's accuracy was above chance (89%, 89%, and 72% correct, respectively, *n*s = 18; binomial tests, *p*s < .048). In addition, the relation between responses and nonreinforced probe trials, nonreinforced baseline trials during criterion learning in Phase 3, and reinforced baseline trials during criterion learning in Phase 3 was not statistically significant for color part-identity, color whole-identity, and shape whole-identity problems, χ^2s (2, *N*s = 93) < 2.19, *p*s > .335, which revealed that response accuracy for novel color part- and whole-identity and novel shape whole-identity problems did not differ from criterion accuracy for familiar color part- and whole-identity and familiar shape whole-identity problems..

Last, Junior's accuracy was not different from chance for nonreinforced probe trials of the novel shape part-identity problems (50% correct; n = 18; binomial test, p =.50); however, the relationship between responses to shape part-identity problems and nonreinforced probe trials, nonreinforced baseline trials during criterion learning in Phase 3, and reinforced baseline trials during criterion learning in Phase 3 was statistically significant, $\chi^2(2, N = 93) = 9.65$, $p = .008$, $V = .32$. The relationship was such that Junior's accuracy was significantly lower for nonreinforced probe trials of the novel shape part-identity problems (50% correct) than nonreinforced baseline trials during criterion learning in Phase 3 (86% correct; $z = 2.41$, $p = .008$) and reinforced baseline trials during criterion learning in Phase 3 (83% correct; $z = 2.83$, $p = .002$), and accuracy did not differ between nonreinforced and reinforced baseline trials during criterion learning in Phase 3 (z = -.25, p = .400), which revealed that response accuracy for novel shape part-identity problems was lower than criterion accuracy for familiar shape part-identity problems.

3.3 Discussion

There were clear differences in the propensity of subjects to learn to select the comparison stimulus that was identical in color to the sample stimulus when color was the only property that differed among stimuli (Experiment 1A). Junior required nearly 4,800 trials before he reached the performance criteria, thus, demonstrating mastery of color whole-identity responding. On the other hand, Madu reached the performance criteria after 500 trials. These differences likely arise from the divergent experimental histories of subjects.

During Experiment 1A, the relation between the sample and correct comparison stimulus (S+) or the relation between the sample and the incorrect comparison stimulus (S-) may have exerted control over the color whole-identity matching behavior of subjects. S+ control can be described by ‘if B is the sample then select B as the comparison’, whereas, S- control can be described by ‘if B is the sample then do not select R or Y as a comparison’. To test for S- control, the positive trained comparison can be replaced by a novel comparison (e.g., B - NOTA, not R) with S- control evidenced if subjects choose the novel untrained comparison. To test for S+ control, the negative trained comparison can be replaced by a novel comparison (e.g., B → B, not NOTA) with S+ control evidenced if subjects choose the familiar, trained positive comparison (Tomonaga, 1993).

Thus, the first 100-trial session of the modified two-choice MTS task with the NOTA comparison stimulus during Experiment 1B (Phase 1 for Madu and Phase 2 for Junior) reflects a test of the controlling relations during color whole-identity matching in Experiment 1A. Junior's behavior illustrated that color whole-identity matching was completely controlled by the S+; his responses were 100% correct during the first 100-trial session of Phase 2 when the NOTA comparison replaced a trained comparison stimulus as the S- stimulus (i.e., NOTA as the incorrect comparison), but were 0% correct when the NOTA comparison replaced a trained comparison stimulus as the S+ comparison (i.e., NOTA as the correct comparison). Similarly, Madu's behavior illustrated that color whole-identity matching was S+ controlled; her responses were 100% correct during the first 100-trials of Phase 1 when the NOTA comparison replaced a trained comparison stimulus as the S- stimulus, but were 40% correct when the NOTA comparison replaced a trained comparison stimulus as the S+ comparison. Indeed, S+ control of the identity matching was described in two chimpanzees called Pan and Panzee for auditory to visual cross-modal identity matching (Beran, 2010; Hashiya & Kojima, 1997), a chimpanzee called Chloe for color to shape arbitrary identity matching (Tomonaga, 1993), and three chimpanzees called Lana, Sherman, and Panzee for photograph to lexigram identity matching (Beran & Washburn, 2002).

With respect to concurrent color whole-identity and -nonidentity responding when color/pattern was still the only property that differed between stimuli during Experiment 1B, the findings showed that Madu learned the task more readily than did Junior. Specifically, Madu required fewer trials to achieve above chance accuracy during a session (200 vs. 600 trials, respectively, for Madu in Phase 1 and Junior in Phase 2) and fewer trials to demonstrate mastery of the task by reaching the performance criteria (500 vs. 800 trials, respectively, for Madu in Phase 1 and Junior in Phase 2).

Moreover, decrements in accuracy occurred when the NOTA stimulus was present as a comparison stimulus during the criterion-level performance of both subjects in Experiment 1B Phase 2. The criterion performance of Madu suggested that she was more likely to make errors when NOTA was present as a comparison, but she did so in a random manner as her accuracy was significantly lower but not different when NOTA was the correct and the incorrect comparison than when NOTA was absent as a comparison. Junior’s criterion performance suggested that he preferred to select the NOTA comparison stimulus when it was present as a comparison as his accuracy was statistically higher but not different when the NOTA comparison was absent and when it was the correct comparison than when NOTA was the incorrect comparison. The aforementioned response patterns likely arose because the simple S+ control that was previously described for identity responding when all comparisons were color stimuli (i.e., if B then select B or if B then do not select R, respectively) could not continue to govern concurrent identity and nonidentity responding because now the S+ was nonidentical to the sample on some trials (i.e., when NOTA was the correct comparison) and on some trials it was identical to the sample (i.e., when NOTA was the incorrect stimulus).

Finally, both subjects required more time to respond to color whole-nonidentity (i.e., when NOTA was the correct comparison) than color whole-identity (i.e., when NOTA was the correct comparison or absent as a comparison) during their criterion performance in Experiment 1B Phase 2. The found response speed difference is likely a product of the unique type of responding that the modified MTS tasks required. In particular, nonidentity judgments during standard nonmatching-to-sample (NMTS) and oddity conditional discrimination tasks require that subjects select the comparison that is nonidentical to the sample in the presence of another comparison that is identical to the sample (i.e., select the nonidentical stimulus if nonidentity exists). The modified MTS task, on the other hand, required that subjects select a particular comparison stimulus that was nonidentical to the sample in the presence of another comparison that was nonidentical to the sample (i.e., select NOTA if nonidentity exists). For this reason, responding successfully to nonidentity with the modified MTS task might have required additional response time.

Turning the discussion to part-identity responding, the accuracy of both subjects for color part-identity problems was above chance during the first session and remained so across all or the majority of sessions during concurrent color and shape part-identity responding in Experiment 2A. On the other hand, the accuracy of both subjects for shape-part-identity problems was not different from chance during the first session and many sessions thereafter. It was not surprising that subjects failed to respond to shape part-identity while successfully responding to color part-identity during the period of early learning following the tasks of Experiment 1. The tasks of Experiment 1 focused on learning to respond to color whole-identity when the shape of stimuli was an irrelevant dimension that remained cue-constant within and between trials (i.e., stimuli were shaped only as squares). The research on learned attention and dimensional relevance in category learning in humans and baboons illustrates that when the cue-to-outcome correspondence shifts at some point during the course of training such that the shifted correspondence has the same relevant, diagnostic, or valid cues as the initial correspondence, then learning the shifted correspondence is relatively fast and easier. If the relevant cues for the shifted correspondence differ from those for the initial correspondence, then learning the shifted correspondence is relatively slow and harder (Fagot, Kruschke, Dépy, & Vauclair, 1998; Kruschke, 1996). Second, there is existing data to support the finding that chimpanzees, orangutans, and monkeys prefer color over shape as the relevant discriminative dimension in matching and oddity tasks (Davis, Leary, Stevens, & Thompson, 1967; Draper, 1965; Garcha & Ettlinger, 1979; King, 1973; Yagi, Shinoda, Shinohara, & Hirata, 1975).

So the orangutans learned attention to color as the relevant dimension may have perseverated because attending to shape as the relevant dimension may be intrinsically more difficult than doing the same for color even though all stimuli began to vary in their shape and the cue-to-outcome correspondence broadened to include shape as the relevant dimension for half of the trials. The former explanation concerning learned attention to dimensional cues would be supported if the order of original learning was reversed such that shape whole-identity tasks were followed by concurrent shape and color part-identity tasks and subjects failed to judge identity for color part-identity problems while successfully judging identity for shape part-identity problems during concurrent color and shape part-identity tasks. This dissertation, however, leaves this issue as a matter for future investigations.

Because the start of the concurrent color and shape part-identity MTS tasks in Experiment 2A occurred six or seven months after the conclusion of color whole-identity tasks in Experiment 1, one cannot conclude that the introduction of shape as a relevant dimension and an irrelevant cue-ambiguous dimension in Experiment 2 (i.e., changing the task to require color and shape part-identity responding) was responsible for the drop from nearly perfect accuracy during criterion color whole-identity learning to 66% and 74% correct for Madu and Junior, respectively, during early color part-identity learning. It is possible that the passage of time was solely responsible for the decline. In any case, the ambiguity as to the source of the decrement during early color part-identity learning for the concurrent color and shape part-identity MTS tasks prohibits concluding that subjects recognized that color was a constituent property of the stimulus even though their matching accuracy for color part-identity problems reached criterion levels. In other words, one cannot yet say that they learned that a part of an object can be identical to the part of another object while other parts of the object are nonidentical.

Completing sessions that required only shape part-identity responding resulted in Madu mastering shape part-identity problems in Experiment 2. Specifically, her accuracy was above what chance would predict during the very first session of shape part-identity problems in Phase 1 when color was always an irrelevant cue-ambiguous dimension that was not associated with reinforcement. On the other hand, Junior required more than just sessions devoted solely to shape part-identity matching to master shape part-identity problems. Indeed, it was only after learning to respond to shape whole-identity that Junior mastered shape part-identity problems. Specifically, Junior continued to respond no differently from chance after completing 1,500 trials of shape part-identity matching in Phase 1. His accuracy rose to above chance levels only after he completed 700 trials of shape whole-identity problems and he went on to demonstrate mastery of shape whole-identity matching by reaching the performance criteria in Phase 2. Following criterion learning for shape whole-identity in Phase 2, his accuracy was above chance during the very first session in which only shape part-identity problems were represented in Phase 3. By the end of shape part-identity learning in Phase 3, his accuracy was no different from Madu's criterion shape part-identity performance. That Junior required shape whole-identity response learning before mastering shape part-identity responding supports the traditional human developmental view that conceptual abilities shift from identity judgments that are based on the global or holistic aspect of objects in early infancy to the ability to judge identity relations that are articulated along the different constituent properties or attributes of objects in late infancy and early childhood (Burns, 1992; Kemler, 1983; Smith, 1984; Smith, 1993; Smith & Heise, 1992).

Although both subjects went on to master shape part-identity problems when they were given in isolation, one still cannot conclude that the subjects were responding to part-identity. The critical test to discern whether subjects recognized identity relations that are articulated along the constituent properties of objects must involve concurrent color and shape part-identity responding. When both color and shape part-identity responding is required and there is no discriminative stimulus to signal what property is associated with reinforcement, one can infer that subjects recognize that an identity relation exists between one, but not another constituent property when they respond accurately to both types of part-identity problems even though it may be more difficult to recognize the part-identity relation between one property.

What happened when subjects were represented with sessions that required concurrent color and shape part-identity matching only one or five days after they reached criterion-level performance with shape part-identity problems in Experiment 2C?

Junior and Madu both judged color and shape part-identity above what chance would predict during the very first session and during all subsequent sessions. In addition, there was not a decline in shape part-identity response accuracy from when they judged shape part-identity in isolation (i.e., when color was always irrelevant cue-ambiguous) to when they were required to do so concurrently with color part-identity. These results indicate that both subjects responded to color and shape as the constituent properties of stimuli.

After mastering sequential color and shape whole-identity MTS tasks and concurrent color and shape part-identity MTS tasks in Experiment 2, both subjects went on to master concurrent color and shape part- and whole-identity MTS tasks during Phase 1 baseline of Experiment 3. The early learning performance of both subjects, however, showed that shape part- and whole-identity accuracy declined while color part- and whole-identity accuracy was unaffected. Both subjects eventually re-established above chance accuracy with shape part- and whole-identity problems, but they were only able to meet a reduced performance criterion of 70% correct for three consecutive days after 1,000 or more trials during concurrent matching with the four identity problem types.

The introduction of nonreinforced baseline trials in Phase 2 disrupted Junior's shape part- and whole-identity responding during the concurrent color and shape part- and whole-identity MTS task. Decreasing the proportion of nonreinforced baseline trials from 36% to 24% of trials in Phase 3 for Junior seemed to aid in the re-establishment of the shape part- and whole-identity responding, but it is possible that simply receiving more trials in Phase 3 stabilized his matching ability.

In any case, the criterion-level performance of Madu during concurrent color and shape part- and whole-identity MTS tasks showed that responding to shape identity was more difficult than responding to color identity. The same pattern was evident in Junior's responding, but it was expressed to a lesser extent. Specifically, Madu's criterion-level reinforced baseline trial accuracy was statistically lower for shape part- and whole-identity problems than color part- and whole-identity problems in Phase 1 and her criterion-level nonreinforced baseline trial accuracy was statistically lower for shape part-identity problems than color part-identity problems in Phase 3. For Junior, criterion-level nonreinforced baseline trial accuracy was statistically lower for shape whole-identity problems than color whole-identity in Phase 3.

An earlier study with four orangutans and one gorilla reported that the average accuracy of subjects across 300 to 360 trials with color and shape part-identity problems did not statistically differ during a concurrent color and shape part- and whole-identity MTS task (Vonk, 2003), however, inspection of subject accuracy during the last 60 trials showed that their accuracy was about 10% to 15% higher for color part-identity problems than shape part-identity problems for the gorilla and 3 of the 4 orangutans.[16] With respect to criterion-level performance, the pattern found in the present dissertation seems to match the pattern found in a previous study with orangutans; that is it is more difficult to respond to shape part-identity than color part-identity. On a related note, there is also existing data to support the finding that color is preferred over shape as the relevant, discriminative dimension in matching and oddity tasks in monkeys (Davis et al., 1967;

---

[16] Whole-identity problems in Vonk (2003) contained an incorrect comparison that differed from the sample and correct comparison in both its color and shape (e.g., red square → red square, not yellow triangle). Whole-identity problems in the current study, however, contained an incorrect comparison that differed from the sample and correct comparison (a) in color, but not in shape or (b) in shape, but not color. For this reason, the results of Vonk on whole-identity responding are not discussed.

Draper, 1965; Garcha & Ettlinger, 1979; Yagi et al., 1975) and chimpanzees and orangutans (Garcha & Ettlinger, 1979; King, 1973).

Additionally, the criterion-level performance of Madu during concurrent color and shape part- and whole-identity MTS tasks showed that responding to part-identity was more difficult than responding to whole-identity. Specifically, Madu's criterion-level nonreinforced baseline trial accuracy was statistically lower for shape part-identity problems than shape whole-identity and for color part-identity problems than color whole-identity in Phase 3. The data from Madu conform to a previous study in monkeys, which revealed that selection of the odd stimulus was more difficult when the irrelevant dimension was cue-ambiguous (i.e., part-identity) than when the irrelevant dimension was cue-constant (i.e., whole-object nonidentity) (Thomas & Frost, 1983). Of final note, the interactive effect of the relevant dimension (color vs. shape) and identity type (part vs. whole) was illustrated in the finding that Madu's criterion-level nonreinforced baseline trial accuracy was not different from chance for shape part-identity problems in Phase 3, whereas, her accuracy was above chance for all other identity types.

Nonhuman primates, thus, can learn that one thing is identical or nonidentical to another thing and that the identity relation can also be applied on a finer level between the constituent properties of those objects. Did this type of responding transfer to a conceptual understanding of part- and whole-identity at the highest level when both relevant and irrelevant dimensions were made novel? Madu's accuracy illustrated full transfer of color part-identity learning as her accuracy was not only above chance, but no different from criterion nonreinforced and reinforced baseline matching for color part-identity problems. She did not exhibit conceptual behavior about shape part- and whole-identity and color whole-identity as her accuracy was not different from chance for the aforementioned problem types during the transfer test. On the other hand, Junior's matching accuracy illustrated full transfer of color part- and whole-identity and shape whole-identity learning as his accuracy was not only above chance, but also no different from criterion nonreinforced and reinforced baseline matching for color part- and whole-identity and shape whole-identity problems. He did not exhibit conceptual behavior about shape part-identity as his accuracy was not different from chance for the aforementioned problem type during the transfer test.

Why did Junior form whole-identity concepts about color and shape, whereas, Madu did not and why did neither subject form a part-identity concept about shape? Three factors may have contributed to these patterns of concept formation failures. First, the extent to which responses generalize to novel exemplars is a direct function of the number of different training stimuli (Katz, Wright, & Bodily, 2007; Lea, 1984; Oden, Thompson, & Premack, 1988; Thompson, 1995; Thompson & Oden, 2000; Wasserman, Kiedinger, & Bhatt, 1988; Wright & Katz, 2006; Zentall et al., 2008). Though concept formation during acquisition may more readily happen when a large number of training exemplars are used, it is also possible that the use of a large number of training exemplars increases the likelihood that any given novel transfer test stimulus will resemble one or more of the familiar training stimuli. The latter scenario may be especially true of color, which is a mixture of light wavelengths, such that providing additional color exemplars may serve as a confounding influence on assessments of concept formation. In any case, the number of color and shape exemplars used during the learning of concurrent part- and whole-identity responding in the present experiments may not have been sufficient to promote more than item-specific learning in Madu for color and shape whole-identity and in both Madu and Junior for shape part-identity.

Second, in the previous chapter I used the idea of levels of concept formation to evaluate the existing literature about color and shape part- and whole-identity concept formation in monkeys, apes, and human infants. The highest level of concept formation reflected an understanding of identity relations for all instances of a dimension regardless of how they were distinctly instantiated in terms of another dimension. In the present study, concept formation was assessed at the highest level by introducing problems in which the instances of both the relevant and irrelevant dimension were novel. If the proposed levels of concept formation are akin to the levels of relational knowledge proposed by Smith (1984) then concept formation should be hardest at the highest level. Concepts may have formed at the lowest and intermediate levels for color and shape whole-identity for Madu and for shape part-identity for both Madu and Junior, but the present experiment did not assess concept formation at this level. Finally, both subjects failing to behave conceptually to shape part-identity may reflect what was generally found in terms of their criterion responding during the concurrent color and shape part- and whole-identity MTS tasks; specifically, that it is not only harder to respond to shape versus color identity and part- versus whole-identity, but that it is also more difficult to form concepts along the same lines.

The literature is rife with examples of apes, monkeys, and human infants learning to respond to color and shape whole-identity, but to a lesser extent illustrates how nonlinguistic apes, monkeys, and human infants learn to respond to part-identity about color and shape. Moreover, sometimes chimpanzees, orangutans, monkeys, and human infants form part- and whole-identity and -nonidentity concepts from their learning (Barros, Galvão, & McIlvane, 2002; Bhatt et al., 2004; Fujita, 1983b; Hayne, Rovee-Collier, & Perris, 1987; Jackson & Pegram, 1970; King, 1973; McMurray & Aslin, 2004; Meyer & Harlow, 1949; Robinson, 1955; Tomonaga, 1995; Truppa et al., 2010; Weinstein, 1945; Young & Harlow, 1943).

Not only did the subjects of the present study learn to respond to identity at the holistic level for certain colors, they also learned to respond to nonidentity at the holistic level for those colors too; moreover, they were able to do the aforementioned concurrently. This experiment, thus, documented the ability of nonlanguage-trained apes to make concurrent identity and nonidentity judgments using a modified MTS format that was akin to the concurrent identity and nonidentity judgments that are required to be made during S/D discrimination tasks. The subjects of the present study also learned to respond concurrently to identity at the holistic and at the dimension-differentiated level and their learning transferred at the highest level to the formation of concepts about color part-identity in both orangutans and to the formation of color and shape whole-identity in one orangutan. This experiment, thus, documented the ability of nonlanguage-trained apes to conceptualize identity relations holistically and between constituent properties of objects.

CHAPTER 4

NUMERICAL COMPETENCE

Animals encounter challenges in their daily lives to which the use of quantitative information would be advantageous. For example, animals may decide whether to visit various food patches based on the amount of food present or to engage in aggressive interactions by determining the number of potential opponents (Gallistel, 1989; Hauser, 1997; Tomasello & Call, 1997; Wynn, 1998a). This area of animal cognition is often called numerical competence or cognition, which is perhaps a misnomer because this line of research is devoted also to understanding how nonverbal organisms process and use information about all types of quantitative attributes (Boysen, 1997; for reviews, see Boysen & Capaldi, 1993; Boysen & Hallberg, 2000; Brannon & Roitman, 2003; Davis & Memmott, 1982; Davis & Pérusse, 1988; Dehaene, 1997; Dehaene, Dehaene-Lambertz, & Cohen, 1998; Gallistel, 1989; Gallistel & Gelman, 2000; and Tomasello & Call, 1997).

4.1. What is Number?

A quantitative attribute is an attribute of which the instances are related to one another both additively and ordinally. Following well-established usage derived from the Aristotelian division of mathematics, specific instances of a quantitative attribute may be divided in two ways: first, continuous quantity or magnitude and second, discrete quantity or multitude (Bell, 1937; Detlefsen, 2005; Michell, 1997).

Under the name of continuous quantity comes what is undifferentiated or unified, in other words, what is continuous. For example, the specific length between point *A* and *B* is a magnitude of the quantity length. As one goes from *A* to *B*, the line continues without break, as it is not made up of discrete units. Other examples of magnitudes are time, area, and volume. Magnitudes of a quantity are not collections of things, they are not denumerable, but they are measurable because they can stand in relations to one another that can be expressed as real numbers. For example, $a + b = c$ is the relation between magnitudes, say specific lengths, that denotes that magnitude c is composed of the discrete parts of magnitudes *a* and *b* and these relations can be expressed as real numbers with each number representing many possible relational compositions, like $4 = 3 + 7$ or $4 = 25 - 21$ (Dooley & Gill, 1977a; Michell, 1997) (Bell, 1937; Detlefsen, 2005; Michell, 1997).

Under the name of discrete quantity are discrete units that are indivisible in the sense that if they are divided the result will not be a unit. For example, half of a person is not a person and half of a table is not a table. Discrete quantity is best represented by number or more specifically, by all positive rational integers (e.g., ten humans and five pencils, but not negative ten humans or negative five pencils). Number should not, however, be confused with numerals, which are the symbols used to represent number for counting, measuring, labeling, and ordering sets of things (Bell, 1937; Detlefsen, 2005; Michell, 1997).

What more can be said about number? Number is an abstraction, a discrimination based on a *single* property of stimuli independent of other properties, derived from our perception of the physical world and conceptualizing quantities (Conant, 1896; Dehaene, 1992; Dooley & Gill, 1977a; Hamilton, 1982; Michell, 1997; Strauss & Curtis, 1984; Thomas, 1988; Tomasello & Call, 1997). Piaget (1941/1965) wrote that "the permutations of the elements in a given set do not change its value. A number is only intelligible if it remains identical with itself, whatever the distribution of the units of which it is composed" (p. 3). Resting on Piaget's framework, in this dissertation, number is defined as the only property of a set that remains invariant when perceptual characteristics of the set's elements change. How this definition applies to the cardinal and ordinal aspects of number will be apparent in the next sections.

Of final note, the use of the term number more often than not invokes thoughts of numerals and counting in the way that humans use number, but the quantitative skills of nonverbal organisms need not involve symbols or formal enumerative processes (Dehaene, 1992; Stevens, 1951; Tomasello & Call, 1997). Experimental paradigms in which an organism's responses are controlled primarily by countable or discriminable (viz. by the experimenter) items are usually termed numerosity or numerical discrimination, judgment, and comparison tasks (Davis & Pérusse, 1988; Stevens, 1951), whereas, the term quantity is used as a general term to describe amounts such that both continuous and discrete quantity may be involved in an organism's responses.

4.2. Types of Number

Tomasello and Call (1997) divide the quantitative skills and abilities that nonhuman animals share with human infants before they develop adult-like abilities into two general types: those that require ordination and those that require cardination. Researchers argue whether human infants acquire ordinality before or after they acquire cardinality, whether both develop simultaneously, and whether ordination always involves cardination and vice versa (Brainerd, 1979). The studies that attempt to understand the developmental sequence, though, typically focus on how linguistic children respond to quantities (Henry, 1976; Kingma & Koops, 1981; Piaget, 1941/1965;

Siegel, 1974) so preverbal abilities remain underexplored. A brief discussion of ordinal number is presented first to provide a framework for thinking about how apes, monkeys, and nonverbal human infants understand the cardinal aspect of number.

4.2.1. Conceptual Understanding about Ordinal Number

The ordinal aspect of number concerns the position of an element in a series with respect to some quantitative property and it identifies certain places in a sequence, answering questions about 'which one' it is in the set. For example, one would say that the horse that ran the fastest held 1st place, the horse that ran the next fastest held 2nd place, and the horse that ran the next fastest held 3rd place, but one could also use names like excellent, good, and poor to rank order elements in a set. The central tendency of an ordinal attribute can be represented by its median and mode, but the mean cannot be defined because an ordinal attribute is not quantitative if the differences between its degrees are not additively structured. For example, the arithmetic mean of the ordinal numbers first and second does not reflect the average finish times of the first and second place horses (Fuson, 1988; Hamilton, 1982; Michell, 1997; Stevens, 1951; Tomasello & Call, 1997; Whitehead & Russell, 1927).

Defining number as the only property of a set that remains invariant when other characteristics of the set's elements change is consistent with the ordinal aspect of number. It means that in a set with three elements, there will be a 1st, 2nd, and 3rd position regardless of if the perceptual properties of the set's elements change. For example, in a set with three objects ordered with respect to length, there is a smallest, middle-sized, and largest object regardless of whether the color, shape, and spatial arrangement of the objects change (Brannon & Terrace, 1998; Fuson, 1988; Hamilton, 1982; Michell, 1997; Piaget, 1941/1965; Stevens, 1951; Tomasello & Call, 1997; Whitehead & Russell, 1927). There are many orderings in a set (but a set has only one size) and finding the object with a particular ordinal position involves matching ordinal numbers one-by-one to objects within the set. Thus, ordinal number concerns the relation of stimuli in a set to other stimuli in the set and that means it can be characterized in terms of relational learning and concept formation (Brainerd, 1979; Brannon & Terrace, 1998; Fuson, 1988; Hamilton, 1982; Michell, 1997; Piaget, 1941/1965; Stevens, 1951; Tomasello & Call, 1997; Whitehead & Russell, 1927).

One way that researchers evaluate how organisms understand the ordinal aspect of number is to analyze generative, emergent performances that are derived from relational learning of stimulus sequences and serial organization (De Lillo, 1996; Green, Stromer, & Mackay, 1993; for review, see Thompson & Oden, 2000; and Tomasello & Call, 1997; Vasconcelos, 2008). Methodologically and conceptually, this is the analysis of stimulus classes based on order derived from relational learning and it is analogous to the analysis of stimulus equivalence as proposed by Sidman (1990, 1997). That is, first, the contingencies that establish production of stimulus sequences and serial behavior can lead to the production of sequences that are not trained explicitly (i.e., stimulus classes based on common ordinal positions). Second, an order relation exists if the relations among stimuli in sequences are characterized by the four properties of an order relation: irreflexivity, asymmetry (also called antisymmetry), transitivity, and connectedness (Green, Stromer, & Mackay, 1993; Hamilton, 1982; Levy, 1979; Stevens, 1951).

4.2.2. Conceptual Understanding about Cardinal Number

Cardinality answers questions about ‘how many’ elements are in the set, in other words, cardination concerns assessments of a set’s size. For example, ‘threeness’ characterizes things like the number of sides of a triangle and the number of leaves of a shamrock so the cardinal number three is the class of all trios. The term numerosity may also be applied to indicate the size of a set, for example, the set’s cardinality or numerosity is three. A set has only one size (but there are many orderings in a set) and that size applies to the whole set, not to an individual member of the set.

The cardinal aspect of number rests solely on the principle of setting things in one-to-one relations or correspondence. Finding one-to-one correspondences between two classes is a way of matching them member for member, with members paired until one or both groups are exhausted. The major use for doing such is to establish, without actually counting, that some class is of the same or a different size or is more or less in size as another set of known size (Bell, 1937; Brannon & Terrace, 1998; Dantzig, 1939; Dehaene, 1992; Fuson, 1988; Hamilton, 1982; Michell, 1997; Stevens, 1951; Tomasello & Call, 1997).

Defining number as the only property of a set that remains invariant when other perceptual characteristics of the set’s elements change is consistent with the definition of cardinality. What this means is that the size of a set is three regardless of whether the set is composed of three objects, three people, or three sounds and regardless of whether the shape, color, spatial arrangement of the elements within the set change (Stevens, 1951; Whitehead & Russell, 1927). Because number is itself an abstraction, it is my contention that an organism’s cardinal number understanding should be evaluated within the framework of concept learning just like ordinal number is typically evaluated within the framework of stimulus equivalence classes, which are a special type of associative concept. Indeed, Thomas (1988) thought that it may be redundant to consider number as its own separate conceptual process. So when I write that an organism understands cardinal number, it is meant that the understanding is at a conceptual level, otherwise, terminology like judging and responding to cardinality or cardinal number is used.

How should one assess if an organism understands the cardinal aspect of number at a conceptual level? The first way involves the successful transfer of discriminative ability established with a training set of numerosities to the same numerosities instantiated in sets that have perceptual properties different (i.e., novel) from those of the training set. In other words, responding in a way that shows that one perceives that the cardinal aspect of number is the only property of a set that remains invariant even when the nonquantitative perceptual characteristics (color, shape, spatial orientation, spatial position, and heterogeneity of elements) of the set change. Similarly, Gallistel (1989, 1993) and Gallistel and Gelman (1992) defined understanding the invariance of the cardinal aspect of number in the face of novel perceptual changes as when animals possessed a number category. Of final note, Douglass (1925) recognized that there is no limit to a concept’s extension or perfection; instead, there are varying degrees to the attainment of a perfect concept because a concept is never complete and its limits boundless; so the number of dimensions that novel transfer stimuli should vary from the training stimuli is debatable.

Returning to the original question about how one assesses cardinal number understanding at a conceptual level, the second manner involves the generalization or transfer of discriminative ability established during training with a set of numerosity stimuli to sets that are composed of novel numerosities. In other words, to infer concept learning about cardinal number, an organism that learns to discriminate the numerosity of one set of stimuli should respond appropriately and without further training when novel numerosities are introduced. The common thread that runs between the first and the second manner of assessing conceptual cardinal number understanding is the ability to generalize the results of learning. The ability to generalize the results of learning is to what a concept about cardinal number, like all concepts, ultimately reduces (Strauss & Curtis, 1984; Thomas, 1988).

The term numerosity is applied in this dissertation only to refer to the number of items, objects, or elements within a set of stimuli when the experimental procedures have provided controls to isolate number from continuous quantity and prevent pattern-based responding. Thus, the next topic for discussion concerns how the aforementioned controls may be accomplished. First, one cannot demonstrate that an organism is responding to cardinal number without constructing sets of items that differ systematically in no way other than number, a task that is difficult because continuous quantity covaries with number. For example, as one adds items to an array, the total area that the items cover in the array increases (i.e., positive correlation) and the average distance between items decreases (i.e., negative or inverse correlation) such that number is confounded with these continuous quantities if controls are not employed. Randomizing the presentation of confounding quantitative properties across trials is one way to control this problem, but other methods are used within the empirical literature to ensure that number is the only property that controlling responding (Tomasello & Call, 1997).

To isolate number from other quantitative properties, at a minimum (a) element area, perimeter length, or contour length and (b) inter-element distance or array density must be controlled in some way during visual discrimination tasks. Element area is a measure of the extent of an item's exposed surface. Element area and array brightness are linearly related so by controlling area one also controls brightness. Element perimeter length is a measure of the distance around the sides of a closed polygon. Perimeter length is linearly related to element area for all polygons (i.e., plane figures that are composed of line segments and are bounded by a closed circuit) so by controlling area one also controls perimeter length. Contour length is the circumference or perimeter of a circle (i.e., a polygon of infinite sides), but controlling for contour length does not control for area because the two measures are not linearly related. Finally, it is cumulative element area, perimeter length, and contour length of all the elements within a set that is usually indexed.

Inter-element distance is a measure of the spread of items within an array. Inter-element distance may be calculated in a variety of ways, for example, as the average distance of each item in the array to the center of the array or as the average distance between each item in the array and every other item in the array. Array density is a measure of the number of items per unit area of the array. The length, width, or both length and width of arrays are changed to manipulate density. I note here that some research papers use the term inter-element distance interchangeably with density, but this may lead to confusion because although the two terms are related they are not the same measure. Density is usually positively correlated with the number of items in arrays because the area of the arrays involved in comparisons typically remains unchanged and equivalent; thus, the confounding influence of inter-element distance usually needs to be controlled. Furthermore, imagine discriminating between two sets of 1 to 5 various sized buttons that are placed on two white saucers (the arrays) that are separated by 2 inches. Does the area of the saucers on which buttons are placed or the spread of buttons on the saucers exert a greater confounding influence on the discrimination of cardinal number? I argue that inter-element distance exerts a much greater confounding influence than does array density.

Additionally, one cannot demonstrate that an organism is responding to cardinal number without varying the spatial location of the elements in the array that instantiate a set's numerosity a sufficient number of times to preclude pattern-based responding.[17] Even so, an argument can be made that the number of novel (viz. from the experimenter's perspective) element spatial positions for an array of fixed size is limited such that by, say the first 1,000 to 2,000 trials, the spatial arrangement of elements is not truly novel. Irrelevant dimension cue constancy and cue ambiguity are measures of *between*-stimulus heterogeneity or variability and the prevention of pattern-based responding requires that the spatial location of elements be an irrelevant cue-ambiguous dimension, meaning it must be a dimension that is not correlated with reinforcement and differs in some noninformative way among all stimuli being compared (Bernstein, 1961; Noble & Thomas, 1985; Steirn & Thomas, 1990; Thomas & Frost, 1983). Other irrelevant dimensions like color, shape, and spatial orientation of elements may be cue-constant in

---

[17] The term array refers to the container or location that bounds or groups a set of stimuli (e.g., the lower half of a computer screen or wells on a Wisconsin General Test Apparatus), whereas, the term set refers to the collection or group of stimuli that are used or presented together.

that a specific instance of the dimension is shared by all numerosity stimuli being compared, cue-ambiguous, or they may reflect some combination of cue constancy and ambiguity.

To illustrate the aforementioned, I use as an example two-choice MTS tasks in which it is a given that numerosity is the relevant dimension (i.e., the dimension correlated with reinforcement) and element spatial position is an irrelevant cue-ambiguous dimension: (a) if two red circles, two red circles, and three red circles serve respectively as the sample, correct, and incorrect stimulus then both color and shape are irrelevant cue-constant dimensions; (b) if two red circles, two red squares, and three red triangles serve respectively as the sample, correct, and incorrect stimulus then shape is a irrelevant cue-ambiguous dimension and color an irrelevant cue-constant dimension; and (c) if two red circles, two blue squares, and three yellow pentagons serve respectively as the sample, correct, and incorrect stimulus then both color and shape are irrelevant cue-ambiguous dimensions.

The next sections detail the empirical literature that employs operant techniques to discover whether apes and monkeys form identity and nonidentity concepts about the cardinal aspect of number.[18] The cardinal number of two sets is said to be identical when all things in the sets can be paired off one-to-one such that after the pairing there are no unpaired things in either set (Bell, 1937; Brannon & Terrace, 1998; Dehaene, 1992; Hamilton, 1982; Michell, 1997; Stevens, 1951; Tomasello & Call, 1997). What is not discussed are studies that involve differential reinforcement values being imposed on

---

[18] There were no empirical studies assessing numerical identity in infants using operant or associative learning procedures.

arbitrary stimuli because differences in conditioned reinforcement can account for subjects behaving in a conceptual manner. For example, reinforcing subjects with four pieces of food for selecting the Arabic numeral four would reinforce that choice response more strongly, induce more salivation and satiation, and elicit stronger affective responses than would reinforcing subjects with two pieces of food for selecting the Arabic numeral two such that different reactions may become associated with the different number symbols and serve as the basis of relational responding (see Brannon & Terrace, 1998; Gillan, Premack, & Woodruff, 1981; and Olthof, Iden, & Roberts, 1997). Moreover, continuous quantity (e.g., volume) covaries with the number of food reinforcers given to subjects unless controls prevent it from doing so, thus, the underlying trained basis of responding to symbols may be number, quantity, or both number and quantity (see Beran, Evans, & Harris, 2008; Olthof & Roberts, 2000).

Like the earlier chapters discussing identity concept formation about color and shape: (a) only empirical studies using operant methods in which identity responding as well as subsequent concept formation could be assessed are discussed, (b) studies that involved two or more relevant dimensions (i.e., numerosity and at least one other dimension), cross-modal and extradimensional concept formation, and symbolic language systems (i.e., lexicons and lexigrams) were excluded, (c) assessment of transfer was limited to early-trial transfer performance (25 or fewer trials of a single problem when differential reinforcement was employed), and (d) baseline and transfer test performance was considered equivalent when their accuracy differed by less than 5% or 6% and transfer test performance at what chance would predict if accuracy differed from chance by less than 15% to 16% when statistical evaluations were absent.[19, 20]

4.2.2.1. Identity in Apes

Evidence in favor of an identity concept about cardinal number was demonstrated in an enculturated chimpanzee called Sheba (Boysen & Berntson, 1989). Using one-, two-, and three-choice MTS procedures, the subject was trained to match homogenous and heterogeneous sample sets of 1 to 3 food items to homogenous comparison sets of 1 to 3 metal discs, Arabic numerals I, II, or III, or combinations of metal discs in one or two comparisons sets and Arabic numerals in one or two comparison sets.[21] Even though element area was positively correlated with the number of food items in homogeneous comparison sets and inter-item distance was negatively correlated with the number of metal discs in homogenous sets, food items were presented in a variety of spatial locations within sample sets and some sample sets were heterogeneous so subjects were prevented from matching samples to comparisons based on inter-item distance or element area. The subject's matching ability fully transferred to the matching of sample sets of familiar Arabic numerals to heterogeneous comparison sets that were composed of 1 to 3 novel items presented in three-choice MTS tasks. From these results, one can infer that the subject formed an identity concept about cardinal number.

---

[19] Data that showed that a concept *did not form* across more than 25 differentially reinforced trials of a single problem were considered for inclusion.
[20] When statistical evaluations were absent, chance probabilities for correct responses were estimated to be .50 for two-choice simple and conditional discriminations, .33 for three-choice simple and conditional discriminations, and so forth unless otherwise indicated in the text.
[21] Roman numerals instead of Arabic numerals are used to make it clear that matching involved symbols, not numerosity.

Detailed in two reports, the behavior of a language-trained chimpanzee called Ai provides support for the formation of an identity concept about cardinal number even though additional training with a larger set of exemplars was necessary for concept formation (Matsuzawa, 1985; Matsuzawa, Asano, Kubota, & Murofushi, 1986). The training procedure used a many-to-one conditional discrimination procedure that required the matching of sample sets of Arabic numerals to sets of objects. The objects used in sets were presented in different spatial locations and orientations; thus, inter-element distance varied randomly and did not covary with the number of objects in sets. Training proceeded in a stepwise manner by adding Arabic numerals and objects such that responding was first established with Arabic numerals I and II and sets of 1 and 2 objects and finally with Arabic numerals from I to V and from 1 to 5 objects. Although the area of objects was positively correlated with the number of objects in sets, using the many-to-one MTS procedure with different objects prevented samples and comparisons from being successfully matched based on area (e.g., 1 pencil → I; 1 brick → I).

Transfer was assessed after the learning criterion was reached after every step by introducing nondifferentially reinforced probe trials of a novel object instantiated in a novel color. Initial probe trial accuracy with novel exemplars was at chance levels when Arabic numerals I to III and 1 and 3 objects were matched, which indicates that learning did not transfer to sets instantiated with novel perceptual features. As training progressed stepwise, matching ability partially transferred. Specifically, Ai obtained above chance probe trial accuracy when matching novel exemplars when the sets were composed of I to IV and I to V Arabic numerals and 1 to 4 and 1 to 5 objects (Boysen & Berntson, 1989).

These data show the formation of an identity concept about cardinal number even though additional training with a larger set of exemplars was necessary for concept formation.

The responses of a language-trained adult chimpanzee named Sarah showed the formation of an identity concept about cardinal number, but the responses of four nonlanguage-trained juvenile chimpanzees did not show concept formation (Woodruff & Premack, 1981). First, Sarah and the four juvenile chimpanzees were trained to match to a criterion using a two-choice MTS task that employed 1 to 4 wooden discs, food items, or metal containers of liquid of various sizes as the sample and comparison stimuli. The sample and comparisons for a problem were taken from the same class (i.e., 4 wooden discs → 4 wooden discs, not 2 wooden discs) so all irrelevant dimensions were cue-constant.[22] The sets of food items and wooden discs were positioned in random spatial arrangements so inter-item distance was roughly controlled for in these two types of sets. However, elements were homogenously sized within sets so the element area and contour length of the sample and correct comparison were identical; thus, number was confounded with continuous quantity during training.

To test for concept formation, subjects were required to match sample sets of 1 to 4 liquid containers to comparison sets of 1 to 4 wooden discs or food items (e.g., 2 liquid containers → 2 wooden discs, not 1 wooden disc) under differential reinforcement. Because matching now involved sets in which the sample and comparisons were taken from different classes, the transfer test may be said to employ task-novel stimuli. Moreover, because matching occurred between sets from different classes, there was no

---

[22] It is likely that subjects were not attending to the amount of liquid in the containers, but attending to the containers as whole objects.

systematic relation between the sample and comparison stimuli's element area, inter-item distances, or contour length, other than number. Sarah's matching ability during training was above what chance would predict at 100%, 86%, and 89% correct for wooden discs, food items, and liquid containers during training. The four juveniles responded correctly in a significant proportion of all trials as well (except for one juvenile with food items), but they were less accurate than Sarah was during training. During the transfer test, Sarah's matching performance was statistically above chance immediately and evinced full concept formation however, the accuracy of the four juveniles was not above chance (Woodruff & Premack, 1981).

The behavior of a chimpanzee called Dennis did not indicate the formation of an identity concept about cardinal number, but represented learning to match specific comparisons to specific sample stimuli (Ferster, 1964; Ferster & Hammer, 1966). Using an arbitrary two-choice MTS procedure, the subject was trained to match to a criterion sample visual arrays of elements in a specific shape and color to light patterns as the comparison stimuli. The training set of sample stimuli contained 1 to 3 or 4 white triangles that were arranged in three fixed spatial arrangements (e.g., in a horizontal row, in a vertical row, and as a triangle). The two comparison sets consisted of three circular lights arranged in a row and located on a panel to the right and left of the sample, with specific patterns created by turning lights on and off. Expressed as binary numbers, 0 for light off and 1 for light on, the comparison stimuli were the six or seven specific light patterns that corresponded to binary numbers 1 to 7 (e.g., 1 white triangle → 001, not 010). Transfer was first tested by changing the shape (e.g., triangles to squares) and color of the sample stimuli's elements. Changing the shape of elements in the sample array resulted in accuracy decreasing to nearly 60% correct, which indicates that the subject learned to match specific comparisons to specific samples. After a novel color was introduced and performance re-stabilized, the sample stimulus' elements were arranged in new spatial arrangements (i.e., the rows of elements were changed to irregularly placed elements) to test for transfer a second time. This change disrupted performance such that accuracy again fell to near chance levels (60% correct). The results indicate again that the subject's behavior was controlled by specific learned associations between samples and comparisons, not the cardinal aspect of number (Ferster, 1964; Ferster & Hammer, 1966).

An experiment with a language-trained chimpanzee called Ai suggests that she learned to respond to cardinal number, but did not develop a conceptual understanding about cardinal number (Murofushi, 1997). The subject was trained to match to a criterion homogenous sets of 1 to 7 red blocks and 1 to 7 red pencils to Arabic numerals I to VII. Even though area was positively correlated with the number of items within sets during training, transfer tests used sets of heterogeneous objects so element area was not confounded with the number of items in sets. Specifically, differentially reinforced probe trials of heterogeneous sets of 1 to 7 red and green blocks and pencils (e.g., 3 green blocks and 1 green pencil as a 4-item set) were inserted in training trials to test for transfer. Ai's accuracy dropped by 35% to 40% (across the first 200 trials) when the sets were first made heterogeneous for familiar colored and shaped objects, which shows that her responding was not conceptually based. In particular, if an organism is able to match a collection of, say, five pencils to the Arabic numeral five, but is not able to match a collection of three pencils and two bricks to the same numeral until a large number of differentially reinforced trials have been given then one can only conclude that associative learning is at play, not a conceptual understanding.

In summary, evidence supporting identity concept formation about cardinal number was found in the behavior of one encultured and two language-trained chimpanzees (Boysen & Berntson, 1989; Matsuzawa, 1985; Matsuzawa et al., 1986; Woodruff & Premack, 1981). On the other hand, one of the aforementioned language-trained chimpanzees and five nonlanguage-trained chimpanzees failed to show an identity concept about cardinal number (Ferster, 1964; Ferster & Hammer, 1966; Murofushi, 1997; Woodruff & Premack, 1981). Together, the findings indicates that apes fail and succeed at forming identity concepts about cardinal number equally often, an interpretation that is contrary to that described by most researchers because most researchers base their conclusions about the numerical abilities of apes on studies that do not isolate number from other quantitative properties. The found failures may indicate that apes do not need a conceptually based understanding of a set's size to function adaptively in domains like foraging and social interactions (Brannon & Roitman, 2003; Davis & Pérusse, 1988) and that they prefer to make their choices based on quantitative properties other than number or in conjunction with number (Beran, Evans, & Harris, 2008) even though such choices may not lead to the most optimal payoff.

An important consideration to keep in mind, though, is whether the expressed inabilities and abilities reflect the general cognitive capacity of apes or are they artifacts of experimental design (Wright, 1992). Specifically, within the taxonomic Superfamily Hominidae, chimpanzees are the only nonhuman species in which the cardinal understanding of number has been investigated. Also, there was variation in the expression of abilities and inabilities within the same subject; specifically, the behavior of the language-trained chimpanzee called Ai showed an understanding of the cardinal aspect of number (Matsuzawa, 1985; Matsuzawa et al., 1986), but in a later study her behavior did not (Murofushi, 1997).

Save one experiment with one language-trained and four nonlanguage-trained chimpanzees (Woodruff & Premack, 1981), all of the experiments involved matching arbitrary and nonarbitrary symbols (i.e., Arabic numerals and light patterns) to numerosities instead of matching numerosities to numerosities. Symbols are used to encompass attributes of their real world referents and this is assumed advantageous in studying conceptual behavior in humans, but it is not yet known precisely how their use influences concept formation in apes. When they are employed with ape subjects, symbols are assumed to capture only the requisite numerical attributes and not capture non-numerical and perceptual features that may trigger an interfering response bias (Boysen, 1997). The two chimpanzees for which Arabic numerals were used to assess cardinal number understanding also received training that involved matching between Arabic numerals and quantities, which makes the aforementioned assumption dubious.

Currently, empirical demonstrations of conceptual behavior about cardinal number are restricted to chimpanzees, concern identity rather than nonidentity, and mostly involve the use of symbols. Furthermore, no study has yet investigated whether apes possess a conceptual understanding of cardinal number as indexed by transfer to novel numerosities. For these reasons, it is important that prospective studies provide new lines of evidence about numerical identity concept formation across the sister ape species.

4.2.2.2. Identity in Monkeys

Three number-experienced rhesus monkeys (Feinstein, Mikulski, and Schroeder) and one number-naïve monkey (Boxer) formed an identity concept about cardinal number even though all subjects failed to do so when element shape was unconfounded with number and the number-naïve subject required additional training trials with a larger set of numerosities to do so when element area was unconfounded with number (Cantlon & Brannon, 2007). Delayed two-choice MTS tasks were given with the elements in sets randomly positioned within arrays so inter-element distance did not covary with the number of elements in sets. Elements that composed a stimulus were homogeneous for element shape, color, and size within a stimulus. In the first experiment, the number of elements within stimuli ranged from 1 to 4 elements.

To test for transfer of learning and number-based responding, after reaching the criterion with training trials in which the number and shape of elements were confounded such that either one could be used to match successfully, all-reinforced probe trials in which number was unconfounded with element shape were inserted within training trials. All four subjects failed to identity match above chance on probe trials in which number was unconfounded with shape. In a second test for transfer of learning and number-based responding, after reaching the criterion with training trials in which the number and area of elements was confounded, all-reinforced probe trials in which number was unconfounded with element area were inserted within training trials. The three number-experienced monkeys identity matched statistically above what chance would predict on probe trials in which number was unconfounded with area, but the number-naïve monkey did not.

In a second experiment, the number of elements within stimuli was supplemented to include 6 and 8 such that 18 additional problems were created and the training and testing procedure of the first experiment with one number-experienced monkey and the number-naïve monkey. Like before, both subjects failed to identity match above what chance would predict when element number and shape were unconfounded. Unlike the first experiment, though, both the number-experienced and number-naïve monkey identity matched statistically above what chance would predict when number and area were unconfounded. Although it is unclear why a concept of numerical identity failed to form when number was unconfounded with shape, it was not surprising that the number-experienced monkeys did not require additional training with an extended set of problems to use number as a basis for matching as they had extensive experimental histories with numerical matching and sequential responding when neither array density nor element color, shape, size, and surface area could be used to consistently solve the tasks (see Cantlon & Brannon, 2006; Jordan & Brannon, 2006).

Two rhesus monkeys (Mikulski and Schroeder) formed an identity concept about cardinal number as indexed by the maintenance of discriminative responding when a novel numerosity was introduced for matching within sets of familiar numerosities (Merritt, Rugani, & Brannon, 2009). Subjects were trained to match using a two-choice MTS procedure with a set of numerosities (1 to 4, 6, 8, and 12). Elements (circles) were presented within a yellow rectangle of fixed size (i.e., the array). The elements that instantiated a numerosity stimulus were homogeneously sized and colored within a stimulus. Sets could be cue-ambiguous or cue-constant for irrelevant dimensions of color and shape. In any case, element color, size, and spatial location varied randomly between and within trials; thus, matching could not be accomplished using element area, contour length, or inter-element distance. Both Mikulski and Schroder matched above what chance would predict during training (75% and 63% correct in order by subject).

To assess transfer of learning, an empty set (0 elements) was introduced into the set of numerosity stimuli. This created 14 semi-novel problems, seven with novel empty set samples (0→ 0, not 1, 2, 3, 4, 6, 8) and seven with familiar samples (1 → 1, not 0; 2 → 2, not 0; 3 → 3, not 0; 4 → 4, not 0; 6 → 6, not 0; 8 → 8, not 0; and 12 → 12, not 0), that were presented within training trials as all-reinforced probe trials. If subjects preferred to select numerosities that were reinforced during training, then they would achieve 100% correct with semi-novel problems with a familiar sample and 0% correct with semi-novel problems with a novel sample such that their overall transfer accuracy would be 50% correct. If a concept about cardinal number formed, then subjects should continue to identity match with both types of semi-novel problems so their overall transfer accuracy should full or partial transfer. Both Mikulski and Schroder's overall transfer accuracy with the semi-novel problems was above what chance would predict and illustrated full and partial transfer (79% and 72% correct in order by subject).

To rule out identity matching based on the array's size and color (e.g., yellow rectangle with no elements → yellow rectangle with no elements, not yellow rectangles with 1 to 4, 6, 8, or 12 elements), additional all-reinforced probe trials of semi-novel problems that had a novel empty set sample were given to subjects. For these trials: (a) the sample stimulus remained unchanged, (i.e., instantiated by the yellow array) and the correct comparison's array was unchanged from or made larger or smaller than its original size or (b) the correct comparison stimulus remained unchanged (i.e., instantiated by the yellow background array) and the sample's array was changed from yellow to one of six colors. Accuracy did not change statistically with any of these changes for any subject; thus, subjects spontaneously recognized identity with a novel numerosity, zero, which illustrates the formation of a concept of cardinal number.

In summary, an identity concept about cardinal number formed in four rhesus monkeys as indexed by transfer of learning to familiar numerosities instantiated with novel perceptual properties in four monkeys (Cantlon & Brannon, 2007) and as indexed by transfer of learning to novel numerosities instantiated with familiar perceptual properties in two monkeys (Merritt, Rugani, & Brannon, 2009). Like with apes, no research study has investigated judgments of numerical nonidentity and subsequent concept formation. In any case, from the available evidence, one can conclude that monkeys form numerical identity concepts without the use of symbols or symbolic language systems.

4.2.2.3. Discussion

A variety of definitions of a number concept have been applied to study the behavior of nonverbal organisms (for review, Davis & Pérusse, 1988). One definition involves the idea of mental representation, the hypothetical internal constructs postulated to be responsible for the observable actions of organisms (Mandler, 1985; Thagard, 1996) such that a concept of number is described as the capacity to integrate symbols from a learned representational numerical system into new emergent relationships (Boysen & Hallberg, 2000).

Other definitions define concept of number in terms of both ordinal and cardinal understanding. First, Dooley and Gill (1977a, 1977b) wrote that the development of a numerical concept of discrete quantity is the recognition of number in a cardinal sense and the recognition of the ordinal interrelationships among individual numbers. Similarly, Piaget believed that concept of number was a synthesis of understanding the cardinal and ordinal aspects of number that formed at the final stage of the development of numerical skills. Specifically, children understand that number may be correctly applied to a collection of entities and these entities can be conceived of as equivalent and therefore grouped into classes and children understand that number may be applied to characterize the relative position of entities in an ordered series and these entities can also be conceived of as equivalent and therefore grouped into classes (Piaget, 1941/1965; see also, Tomasello & Call, 1997; Zimiles, 1963).

Still others define concept of number in terms of arithmetic and mathematics. Gallistel (1989, 1993) and Gallistel and Gelman (1992) wrote that a concept of number is demonstrated when one shows that nonhuman animals perform operations that are isomorphic to some or all of the arithmetic operations that define the number system (i.e., addition, subtraction, multiplication, and division) and the mathematical relations of equality and inequality (e.g., $>$, $<$, $=$, and $\neq$) with the representatives of numerosity. Yet another definition focuses on cross-modality. Dehaene (1997) wrote that the ability to generalize across different modalities of perception or action is an important component of what we call the number concept. For example, recognizing that events such as pressing a lever twice, hearing two sounds, or eating two seeds are instances of the number two, demonstrates a concept of number. Some emphasize both mathematics and cross-modality even though there is considerable latitude in how stringent a criterion one applies. Finally, Davis and Pérusse, (1988) define the number concept in terms of counting, a mathematical action or activity, and the capacity of transfer across sensory modalities and methodology (e.g., transfer of perceived number into performed number or transfer of number discrimination from simultaneous to sequential methods).

Douglass (1925) recognized that there is no limit to a concept's extension or perfection because a concept is never complete and its limits boundless; instead, there are varying degrees to the attainment of a perfect concept. Thus, the discussed definitions of number concepts are all possible components of a fully developed number concept (Strauss & Curtis, 1984). The perfect number concept likely involves generalization based on both the cardinal and ordinal aspects of number and the ability to perform arithmetic and mathematic operations on cardinal and ordinal number. With that said, what processes are proposed to account for the ability of nonverbal organisms to put sets of objects or events in one-to-one correspondence with other sets to judge cardinal number identity, which then forms the basis of forming concepts of identity about cardinal number?

The object-file model is a prominent model proposed to account for numerical processing in humans (Kahneman, Treisman, & Gibbs, 1992). This model proposes that the visual system opens object files, temporary placeholders or tokens, for each relevant object when individuals scan an array of objects. At first, the tokens carry no information about the perceptual characteristics (spatial, temporal, and physical) of objects; instead, these details are filled in later by attentional processes. The system attempts to place currently perceived tokens (e.g., during testing with novel and familiar sets) in one-to-one correspondence with tokens from preceding scenes (e.g., arrays presenting during habituation) with mismatches resulting in preferential visual fixation. The object file model has been recently reformulated in terms of a working memory system called parallel individuation.

Parallel individuation can hold individuals (or objects) that are represented by unique mental symbols in parallel and support identity-nonidentity and more-less comparisons between the working memory set of individuals and a visible set of set to determine one-to-one correspondence. Infants are limited to working memory modules of at least two sets of three or fewer individuals because of hard capacity limits, whereas, working memory modules may create sets of four or fewer individuals for adults (Le Corre & Carey, 2007). Based on empirical evidence on the limits of parallel visual individualization of objects, there is assumed to be a strict limit on the number of objects that can be simultaneously tracked, no more four (Carey, 2001; Hauser, 1997; Simon, 1997). The object file model predicts that nonverbal organisms should fail with numerosity comparisons involving sets with more than four items. Developmentally, object-file models and parallel individuation support discontinuity in the mode of numerical processing as infants must gain additional systems to support processing with larger sets. Finally, the object-file model is limited to comparisons in the visual system and has no built-in mechanism to account for ordinal relations (Brannon, 2002; Brannon & Roitman, 2003).

The accumulator model is the other prominent model proposed to account for counting and timing in animals (Church & Meck, 1984; Gibbon, 1977; Meck & Church, 1983). An animal is said to be counting if the number of events, independently of their duration, serves as a discriminative stimulus and said to be timing if the duration of an event, independent of the number of events, serves as a discriminative stimulus. In this model, a pacemaker puts out energy pulse of relatively fixed duration and a switch passes these pulses into an accumulator. The pacemaker-switch-accumulator system functions as a counter if the switch is operated in the event mode in which it closes for a relatively fixed interval of time to gate pulses into the accumulator at a relatively constant rate. Thus, countable quantity is encoded as magnitudes because the accumulator fills up in equal increments for each entity to be counted; for example, ten entities fills the accumulator up five times as much as two entities does.

The value in the accumulator may be passed onto working memory such that the current accumulator value is compared to a remembered accumulator value (from a time of reinforcement of a previous response) stored in reference memory. In the final step, the decision process occurs, that is, a response rule determines the response based on the comparison that was made (e.g., push the left button or push the right button). Based on experience, the animal learns the accumulator value associated with reinforcement because this information is stored in reference memory. The existence of numerous separate accumulators allows identity-nonidentity comparisons to be made by detecting one-to-one correspondence through simultaneous decrements in two accumulators by one increment until one or both are empty. If both accumulators empty at the same time, then they encode the same number of elements. If not, then the two accumulators encode a different number of elements, with lesser numerosity encoded by the first to become empty (Gallistel & Gelman, 2000; Wynn, 1998a, 1998b).

In the pacemaker-switch-accumulator system, the source of variance that receives the most attention involves the drift in the time between pulses that are generated by the pacemaker. The time between pulses is assumed to be fixed within any trial, but vary normally around a mean between trials. The pacemaker-switch-accumulator counting system, is a scalar (variance in the accumulator increases proportionally to the mean number of counts) counting system (Church & Meck, 1984; Gibbon, 1977; Gibbon & Meck, 1984; Meck & Church, 1983). The second source of variance that has received attention involves the comparator. Responses are theorized to be based on whether the current value in the accumulator is closer to the value in reference memory for a reinforced response. The measure of closeness is the ratio between the current accumulator value and the reference memory value, which corresponds to an acceptance region around the remembered positive value. Together, the ratio comparator combined with scalar counting in the accumulator model should result in accuracy conforming to Weber's law for the discriminability of numerosity comparisons (Gallistel & Gelman, 1992; Gallistel & Gelman, 2000; Gibbon, 1977).

Weber’s law is a psychophysical rule that is usually characterized as constant discriminability with a constant ratio of increment in stimulation to a standard stimulus (Church & Meck, 1984; Gibbon, 1977; Gibbon & Meck, 1984; Meck & Church, 1983). If numerosities are encoded and processed as magnitudes then the more nearly equal two numerosities are, the harder it should be to determine which is larger or which is smaller and the larger two numerosities are at a given difference between them, the harder it should be to determine which is the larger or the smaller. These patterns have been named the numerical size (or magnitude), distance (or difference), and ratio (or disparity) effect. The numerical size effect is the finding that discrimination ability declines (and response latency increases) as the numerical magnitude (i.e., arithmetic sum) of compared numerosities increases for any given numerical distance. The numerical distance effect is the finding that discrimination ability declines (and response latency increases) as the numerical difference (i.e., the arithmetic difference) of compared numerosities declines. The numerical ratio effect combines the numerical distance and size effects as it is the finding that discrimination ability declines (and response latency increases) when the numerical ratio (i.e., the smaller numerosity divided by the larger numerosity) between compared numerosities approach a value of one (Dehaene & Changeux, 1993; Dehaene, Dehaene-Lambertz, & Cohen, 1998).

What do the data from the empirical studies about concept formation reveal about the processes that underlie conceptual judgments about cardinal number? The findings in apes and monkeys show that cardinality is not encoded exactly even at the conceptual level, but instead it is encoded as magnitudes with scalar variability as proposed in the accumulator model (Gallistel & Gelman, 2000). In apes, only a single study reports accuracy and response latency in terms of the numerosity of sets (Matsuzawa, 1985; Matsuzawa et al., 1986). This report indicated that the training and transfer performance of one language-trained chimpanzee called Ai showed the numerical distance and numerical size effects. Most of her errors involved selecting the neighboring Arabic numeral comparison (94% and 96% of errors during training and transfer testing) and of these errors, nearly half involved the largest two Arabic numerals and object sets (48% and 47% during training and transfer testing). Also, her response latency increased as the number of objects in the sample set increased during training, but this trend was more apparent during the initial training sessions than the final training sessions. Object-file models would predict a failure to identity match (and thus form an identity concept) for comparisons that involved more than three or four items; instead, Ai successfully learned to match from Arabic numerals I to V to sets of 1 to 5 objects. Together, these aforementioned patterns are consonant with the predictions of the accumulator model.

In monkeys, the findings indicate that they are more accurate and sometimes faster in making correct choices as the numerical distance between comparisons increase and as the ratio between comparisons move away from a numerical value of one. The probability of making a correct match increased as the ratio between comparisons decreased in the four rhesus monkeys during transfer testing with familiar numerosities that were instantiated with novel perceptual properties (Cantlon & Brannon, 2007). Also, matching accuracy declined and latency increased as the numerical distance between comparisons decreased during transfer testing with familiar and semi-novel sample and comparisons in one of the aforementioned monkeys (Merritt, Rugani, & Brannon, 2009). Again, object-file models would predict a failure to identity match (and thus form an identity concept) for comparisons that involved more than three or four items; instead, successful matching and concept formation was not limited to sets composed of four or less elements in all monkeys. Together, these finding are consistent with the hypothesis that numerical comparison errors are rooted in the noisiness of numerical processing as proposed by the accumulator model.

To end the discussion, I note that to date no study has systematically explored the sole effect of irrelevant dimension cue ambiguity and constancy, which are measures of between-stimulus heterogeneity or variability, on numerical identity judgments and concept formation in apes or monkeys. By sole effect, it is meant that within-stimulus heterogeneity or variability is fixed at the lowest level such that the instances of each irrelevant dimension other than element spatial position are uniform or homogenous within a stimulus instead (e.g., the sample is instantiated as four red circles and not as two red circles, one green square, and one yellow circle). Because cardinal number represents how many discrete things exist in a set regardless of the manner in which those discrete things are instantiated, it is a constituent property of a collection of objects. It seems reasonable then to assume that it would be harder to respond to numerical identity when irrelevant dimensions are cue-ambiguous than when they are cue-constant because earlier experiments found the same pattern with respect to judgments of part-identity about non-numerical dimensions (Bernstein, 1961; Noble & Thomas, 1985; Steirn & Thomas, 1990; Strong, Drash, & Hedges, 1968; Thomas & Frost, 1983). Further, it remains unanswered whether apes can learn to judge numerical identity when all irrelevant dimensions are cue-ambiguous and within-stimulus heterogeneity is fixed at its lowest point, which is presumably when it would be most difficult to do so, *before* they learn to respond to numerical identity when irrelevant dimensions are cue-constant, which is presumably when it would be easiest to do so. Additional data is necessary to provide insights about the effect of between-stimulus variability on numerical identity judgments and concept formation.

CHAPTER 5

EXPERIMENTAL INVESTIGATION OF CARDINAL NUMBER IDENTITY-NONIDENTITY RESPONDING AND CONCEPT FORMATION

The objective of Experiment 4 was to establish concurrent identity and nonidentity responding to cardinal number when both irrelevant dimensions (element color and shape) were cue-ambiguous, which was presumably when accurate conditional responding would be the most difficult. The hypothesis was that subjects would learn to respond to numerical identity when element color and element shape were both cue-ambiguous irrelevant dimensions.

The objective of Experiment 5 was to establish identity responding to cardinal number when both irrelevant dimensions were cue-ambiguous before doing the same when at least one and then both irrelevant dimensions were cue-constant. The hypothesis was that subjects would learn to respond to numerical identity, at the very least, when element color and element shape were both cue-constant irrelevant dimensions; further, I predicted that subject responses would be affected by the numerical distance and numerical size of comparisons.

Finally, the objective of Experiment 6 was to assess transfer of the established conditional discrimination to novel transfer problems in which novel and familiar numerosities were instantiated with familiar element colors and shapes or with novel element colors and shapes and the irrelevant dimensions of element color and shape were cue-constant, one cue-ambiguous and the other cue-constant, or both cue ambiguous. The prediction was that learning would transfer to some of the novel transfer problems.

5.1 Method

5.1.1 Subjects, Housing, Apparatus, and Materials

The subjects, housing, apparatus, and materials were as described for Experiments 1, 2, and 3 in Chapter 3.

5.1.2 Visual and Auditory Stimuli

Visual and auditory stimuli were the three shapes (rectangle, pentagon, and circle), three colors (brown, red, and yellow), one pattern (black and white), and three sounds (sample-touched, correct-response, and incorrect-response tones) described in Experiments 1, 2, and 3 in Chapter 3. In addition, the following three novel shapes and three novel colors were employed: crosses, cylinders, and triangles; and blue (RGB: 0, 255, 255), green (RGB: 0, 255, 0), and white (RGB: 255, 255, 255).

5.1.3 General Procedure

The procedure was as described for Experiments 1, 2, and 3 in Chapter 3. Additionally, the experimenter delivered reinforcement according to both a continuous and partial reinforcement schedule. Subjects advanced to a different experiment part (A, B, and C) or phase (1, 2, 3, etc.) when they meet a specified performance criteria.

5.1.3.1 Experiment 4: Concurrent Numerical Identity-Nonidentity Responding with Irrelevant Dimensions Cue-Ambiguous

Subjects participated in two-choice matching-to-sample (MTS) tasks for one to two sessions per day: Junior for 45 days between March 2, 2010 and June 13, 2010 and

Madu for 47 days between March 2, 2010 and June 5, 2010.[23] The purpose of this experiment was to establish concurrent numerical identity and nonidentity responding, when presumably it was hardest to do so, with both irrelevant dimensions cue-ambiguous.

A numerosity stimulus was composed of a number of elements placed within a 500 x 350 pixel, grey rectangle that served as the array. The number of elements within the array defined the stimulus' numerosity. The numerosity of stimuli was 2, 4, or 6 elements. The numerosity of stimuli was always the relevant dimension (i.e., correlated with reinforcement). Element area (size) was the cumulative area of all elements in the array and it varied according to five sizes: 10,000, 12,000, 14,000, 16,000, or 18,000 pixels. Element area was homogenous or uniform within a stimulus, but differed between stimuli. Inter-element distance was the average of all pairwise distances as measured from the closest edge between element pairs and it varied according to five distance levels: 25, 35, 45, 55, and 65 pixels. Elements were pseudorandomly placed at different spatial locations in the array with the restriction that they be at least 10 pixels apart. The sample, correct, and incorrect comparison stimulus always had different, randomly determined element areas, inter-element distances, and element spatial locations. Thus, the sample's element area and inter-element distance was sometimes greater and sometimes less than one or both comparison stimuli, which prevented numerosity from covarying with inter-element distance and element area.[24]

---

[23] Before participating, subjects completed Experiment 1, which is detailed in Chapter 3.
[24] Junior and Madu received respectively 60 and 75 trials in which the correct and incorrect comparison stimulus had the same element area because of experimenter error.

The elements that instantiated a stimulus' numerosity were colored yellow (Y), red (R), and brown (B) and they were shaped as rectangles, circles, or pentagons. Element shape and color were uniform within a stimulus; in other words, within-stimulus variability was always homogeneous. The shape and color of the elements that instantiated a stimulus were irrelevant dimensions that were cue-ambiguous such that the color and shape of elements within a numerosity stimulus differed from each other. In addition to the numerosity stimuli, a black and white patterned rectangle (500 x 350 pixels) was added to the set of comparison stimuli to serve as the none-of-the-above comparison stimulus (NOTA).

There were nine numerical identity problems for which a correct response was to select the comparison stimulus that had the same number of elements as the sample stimulus and an incorrect response was to select the comparison stimulus with a different number of elements than the sample or to select the NOTA comparison stimulus: 2 → 2, not 4, 6, or NOTA; 4 → 4, not 2, 6, or NOTA; and 6 → 6, not 2, 4, or NOTA. In particular, the sample stimulus was presented with one numerosity comparison stimulus that had the same number of elements as the sample and (a) another numerosity comparison stimulus that had a different number of elements as the sample (the NOTA absent comparison pair type) or (b) the NOTA comparison stimulus (the NOTA incorrect

---

Still, the sample stimulus could not be matched to the comparison based on shared area so these trials were not removed from the analyses except where otherwise noted.

comparison pair type). Figure 5.1.1 shows an example identity problem in which NOTA is the incorrect comparison and NOTA is absent in the comparison.[25]

There were six numerical nonidentity problems for which a correct response was to select the NOTA comparison stimulus and an incorrect response was to select the comparison stimulus with a different number of elements than the sample: 2 → NOTA, not 4 or 6; 4 → NOTA, not 2 or 6; and 6 → NOTA, not 2 or 4. In particular, the sample stimulus was presented with one numerosity stimulus that had a different number of elements as the sample and with the NOTA comparison stimulus (the NOTA correct comparison stimulus pair type). Figure 5.1.1 shows an example nonidentity problem in which NOTA is the correct comparison.

The number of times that each numerosity, color, and shape served as the sample stimulus was balanced within sessions and the presentation order for the sample's numerosity, color, and shape was randomized within sessions. The number of times the NOTA stimulus served as a comparison stimulus was also balanced across sessions. The spatial location of elements in the array was trial-unique for each numerosity stimulus for three-fourths of all sessions given; in other words, some sessions utilized earlier sets. Only two spatial positions for the comparison stimuli (the left and right positions) were utilized and the spatial position of the correct and incorrect stimulus was randomized across trials. Trials were separated by a 2-s ITI. The performance criteria for mastery of tasks was to achieve above chance accuracy (i.e., 59% correct, $n = 100$, $p < .05$) for three consecutive sessions with a minimum of five sessions completed.

---

[25] In all figures, letters are shown within elements only to aid readability. Additionally, all figures depict problems following a touch to the sample (i.e., sample and comparison stimuli are displayed simultaneously).

Reinforcement was delivered on a continuous schedule after every correct response (CRF) unless otherwise noted that it was delivered on a variable ratio intermittent reinforcement schedule after an average of three correct responses (VR-3). The two schedules were used because previous research has shown that reinforcement schedules affect the accuracy of conditional discrimination during early learning in monkeys (Ferster, 1960; Fujita, 1985).

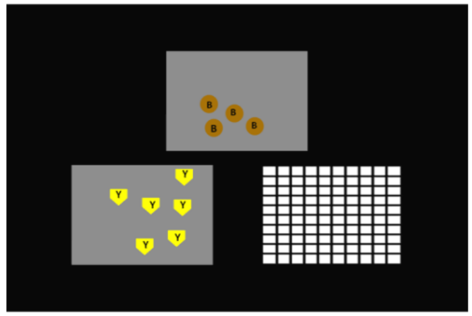
5.1.3.1.1 Color and Shape Cue-Ambiguous (Phase 1)

The purpose of this phase was to establish concurrent numerical identity and nonidentity responding under CRF. The color and shape of the sample stimulus' elements, the correct comparison stimulus' elements, and the correct comparison stimulus' elements differed from each other. Subjects advanced to Phase 2 when they did not meet the performance criteria after completing 4,000 trials.

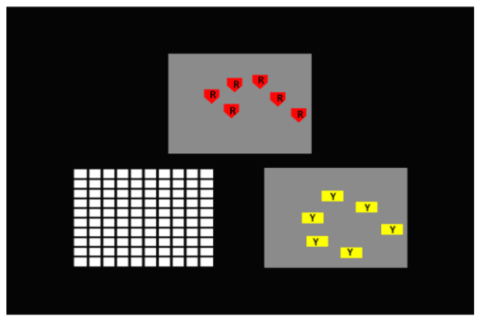
5.1.3.1.2 Color and Shape Cue-Ambiguous under VR-3 (Phase 2)

This purpose of this phase was to establish concurrent numerical identity and nonidentity responding under VR-3. Trials were identical to Phase 1 except that reinforcement was delivered under VR-3. Subjects advanced to Experiment 5 when they did not meet the performance criteria after completing 1,000 trials.

A



B



C

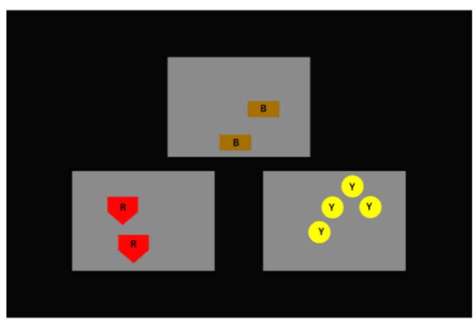


Figure 5.1.1. Example problems during concurrent numerical identity and nonidentity responding with both irrelevant dimensions cue-ambiguous in Experiment 4. (A): Nonidentity problem with NOTA as the correct comparison (4 brown circles [12,000 pixel element area, 25 pixel inter-element distance] → NOTA, not 6 yellow pentagons [16,000 pixel element area, 55 pixel inter-element distance]). (B): Identity problem with NOTA as the incorrect comparison (6 red pentagons [12,000 pixel element area, 35 pixel inter-element distance] → 6 yellow rectangles [16,000 pixel element area, 55 pixel inter-element distance], not NOTA). (C): Identity problem with NOTA absent (2 brown rectangles [12,000 pixel element area, 65 pixel inter-element distance] → 2 red pentagons [16,000 pixel element area, 35 pixel inter-element distance], not 4 yellow circles [18,000 pixel element area, 25 pixel inter-element distance]).

5.1.3.2 Experiment 5A: Numerical Identity Responding with Irrelevant Dimensions Cue-Ambiguous

Subjects participated in two-choice MTS tasks for one to three sessions per day: Junior for 16 and 19 days respectively between May 14, 2010 and June 5, 2010 and November 18, 2010 and December 5, 2010 and Madu for 16 days and 18 days respectively between May 14, 2010 and June 5, 2010 and October 21, 2010 and November 9, 2010.[26]

The purpose of this part of the experiment was to establish numerical identity responding, when presumably it was hardest to do so, with both irrelevant dimensions cue-ambiguous after failures to establish concurrent identity and nonidentity responding. The procedures of Experiment 4 were replicated except the NOTA stimulus (i.e., the black and white patterned rectangle) was not used; thus, there were six identity problems: 2 → 2, not 4 or 6; 4 → 4, not 2 or 6; and 6 → 6, not 2 or 4.

5.1.3.2.1 Color and Shape Cue-Ambiguous under VR-3 (Phase 1)

Reinforcement was delivered on a VR-3 schedule. Subjects advanced to Phase 2 when they did not meet the performance criteria after completing 2,500 trials.

5.1.3.2.2 Color and Shape Cue-Ambiguous (Phase 2)

Trials were identical to Phase 1 except that CRF was used. Subjects advanced to Phase 3 when they did not meet the performance criteria after completing 2,000 trials.

5.1.3.2.3 Color and Shape Cue-Ambiguous for Sample Numerosity of 4 (Phase 3)

---

[26] In the interim between the two periods, subjects completed Experiment 2A, 2B, and 2C, which are detailed in Chapter 3.

The purpose of this phase was to establish numerical identity responding by reducing the number of problems to two. Trials were identical to Phase 2 except that only problems with a sample numerosity of four were employed. Subjects advanced to Phase 4 when they did not meet the performance criteria after completing 500 trials.

5.1.3.2.4 Color and Shape Cue-Ambiguous for Sample Numerosity of 2 (Phase 4)

The purpose of this phase was to establish numerical identity responding by reducing the number of problems to two. Trials were identical to Phase 2 except that only problems with a sample numerosity of two were employed. Subjects advanced to Phase 5 when they did not meet the performance criteria after completing 500 trials.

5.1.3.2.5 Color and Shape Cue-Ambiguous for Sample Numerosity of 6 (Phase 5)

The purpose of this phase was to establish numerical identity responding by reducing the number of problems to two. Trials were identical to Phase 2 except that only problems with a sample numerosity of six were employed. Subjects advanced to Experiment 5B when they did not meet the performance criteria after completing 500 trials.

<u>5.1.3.3 Experiment 5B: Numerical Identity Responding with Irrelevant Dimensions Cue-Ambiguous and Cue-Constant</u>

Subjects participated in two-choice MTS tasks for one to two sessions per day: Junior for 7 days between December 6 and 13, 2010 and Madu for 6 days between November 11 and 17, 2010.

This purpose of this part of the experiment was to establish numerical identity responding, at presumably the intermediate level of difficulty, when one irrelevant dimension was cue-constant and the other cue-ambiguous. The procedures of Experiment 4 were replicated except that one irrelevant dimension was cue-ambiguous and the other was cue-constant and the NOTA comparison stimulus was not used; thus, there were six identity problems (2 → 2, not 4 or 6; 4 → 4, not 2 or 6; and 6 → 6, not 2 or 4).

5.1.3.3.1 Shape Cue-Ambiguous and Color Cue-Constant (Phase 1)

The color of a numerosity stimulus' elements was the same among the sample, correct comparison, and incorrect comparison stimuli while the shape of a numerosity stimulus' elements differed among the sample, correct comparison, and incorrect comparison stimuli. The left panel of Figure 5.1.2 shows an example problem. Subjects advanced to Phase 2 when they did not meet the performance criteria after completing 500 trials.

5.1.3.3.2 Color Cue-Ambiguous and Shape Cue-Constant (Phase 2)

The color of a numerosity stimulus' elements differed among the sample, correct comparison, and incorrect comparison stimuli while the shape of a numerosity stimulus' elements was the same among the sample, correct comparison, and incorrect comparison stimuli. The right panel of Figure 5.1.2 shows an example problem. Subjects advanced to Experiment 5C when they did not meet the performance criteria after completing 500 trials.

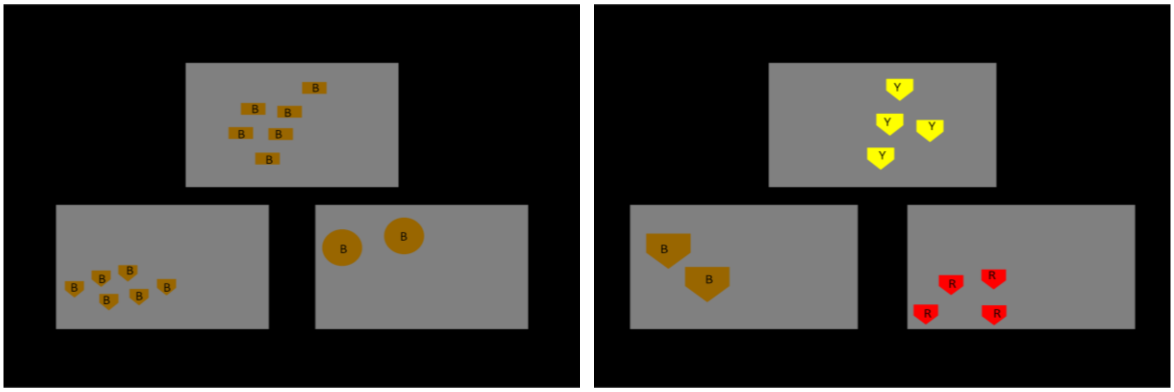


Figure 5.1.2. Example problems with irrelevant cue-ambiguous and cue-constant dimensions for Experiment 5B. The left panel illustrated color cue-constant and shape cue-ambiguous irrelevant dimensions (6 brown rectangles [12,000 pixel element area, 35 pixel inter-element distance] → 6 brown pentagons [10,000 pixel element area, 25 pixel inter-element distance], not 2 brown circles [16,000 pixel element area, 65 pixel inter-element distance]). The right panel illustrates shape cue-constant and color cue-ambiguous irrelevant dimensions (4 yellow pentagons [12,000 pixel element area, 35 pixel inter-element distance] → 4 red pentagons [10,000 pixel element area, 45 pixel inter-element distance], not 2 brown pentagons [16,000 pixel element area, 25 pixel inter-element distance]).

<u>5.1.3.4 Experiment 5C: Numerical Identity Responding with Irrelevant Dimensions Cue-Constant</u>

Subjects participated in two-choice MTS tasks for one to four sessions per day: Junior for 5 days between December 14 and 21, 2010 (only Phase 1) and Madu for 23 days between November 17, 2010 and December 15, 2010. This purpose of this part of the experiment was to establish numerical identity responding, when it presumably was easiest to do so, with both irrelevant dimensions cue-constant. The procedures of Experiment 4 were replicated except that color and shape were both cue-constant and the

NOTA stimulus was not used; thus, there were six identity problems (2 → 2, not 4 or 6; 4 → 4, not 2 or 6; and 6 → 6, not 2 or 4). Figure 5.1.3 shows an example problem.

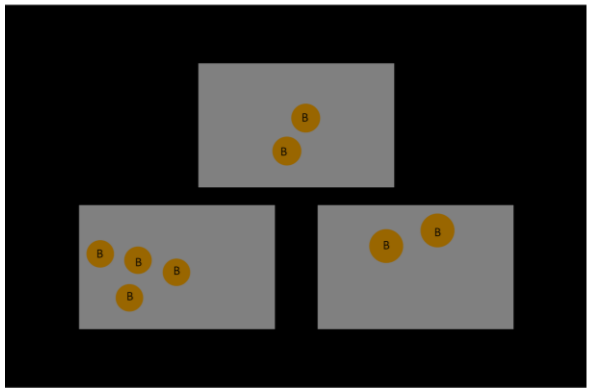


Figure 5.1.3. An example problem with both irrelevant dimensions cue-constant for Experiment 5C. Color cue-constant and shape cue-ambiguous is illustrated (2 brown circles [10,000 pixel element area, 25 pixel inter-element distance] → 2 brown circles [14,000 pixel element area, 55 pixel inter-element distance], not 4 brown circles [18,000 pixel element area, 35 pixel inter-element distance]).

5.1.3.4.1 Color and Shape Cue-Constant (Phase 1)

The color and shape of a numerosity stimulus' elements was the same among the sample, correct comparison, and incorrect comparison stimuli. Except for one 14-trial session, sessions contained 100 trials. After completing 914 trials, Junior failed to respond for 30 minutes for three consecutive days so he was terminated from Experiment 4 and he did not advance to Experiment 6. Madu advanced to Phase 2 when she met the performance criteria.

5.1.3.4.2 Color and Shape Cue-Constant for Sample Numerosity of 4 (Phase 2)

The purpose of this phase was to establish numerical identity responding by reducing the number of problems to two. Trials were identical to Phase 1 except that only problems with a sample stimulus numerosity of four were employed. The subject advanced to Phase 3 when she met the performance criteria.

5.1.3.4.3 Color and Shape Cue-Constant for Sample Numerosity of 2 (Phase 3)

The purpose of this phase was to establish numerical identity responding by reducing the number of problems to two. Trials were identical to Phase 1 except that only problems with a sample stimulus numerosity of two were employed. The subject advanced to Phase 4 when she met the performance criteria.

5.1.3.4.4 Color and Shape Cue-Constant for Sample Numerosity of 6 (Phase 4)

The purpose of this phase was to establish numerical identity responding by reducing the number of problems to two. Trials were identical to Phase 1 except that only problems with a sample stimulus numerosity of were employed. The subject advanced to Phase 5 when she met the performance criteria.

5.1.3.4.5 Color and Shape Cue-Constant (Phase 5)

This phase was a replication of Phase 1 to allow the subject a second attempt to reach the performance criteria with all six numerical identity problems. When the subject met the performance criteria, she was given 15 additional sessions to allow her performance to stabilize before she finished Experiment 5.

<u>5.1.3.5 Experiment 6: Numerical Identity Responding and Test of Transfer</u>

Madu participated in two-choice MTS tasks for two to eight sessions per day for 9 days between December 18 and 29, 2010. The purpose of this experiment was to assess transfer of numerical identity responding from a small set of familiar numerosities instantiated with familiar element colors and shapes when both irrelevant dimensions were cue-constant (i.e., the stimuli from Experiment 5C) to novel and familiar numerosities instantiated with novel and familiar element colors and shapes when both irrelevant dimensions were cue-constant, both cue-ambiguous, or one cue-ambiguous and one cue-constant.

5.1.3.5.1 Reinforced and Nonreinforced (36%) Baseline (Phase 1)

The purpose of this phase was to familiarize subjects to nonreinforced trials; thus, the familiar, trained numerical identity problems of Experiment 5C when both irrelevant dimensions were cue-constant were presented to subjects as two types of trials. Trials of the familiar, trained numerosities from Experiment 5C are called baseline trials. Nonreinforced baseline trials were trials in which responses were differentially reinforced like in all Phases of Experiment 5C and nonreinforced baseline trials were trials that were not followed by the correct- or incorrect-response tone and food reinforcement regardless of the subject's response. Of the 100 trials in a session, 36% were nonreinforced baseline trials and the remaining 64% were reinforced baseline trials. Reinforced and nonreinforced baseline trials were pseudorandomly mixed together within a session in a way that prevented more than three consecutive nonreinforced trials. After meeting the performance criteria, the subject received 15 additional sessions to ensure stable performance before advancing to the transfer test of Phase 2.

5.1.3.5.2 Transfer Test (Phase 2)

This phase replicated Phase 1 except for the following. The percentage of nonreinforced baseline trials was reduced to 18% of trials within a session while the percentage of reinforced baseline trials remained at 64%. The remaining 18% of trials within a session were nonreinforced probe trials. The three trial types (i.e., reinforced baseline, nonreinforced baseline, and nonreinforced probe trials) were pseudorandomly mixed together in a way that prevented more than three consecutive nonreinforced trials.

Nonreinforced probe trials presented problems in which (a) the numerosity of stimuli was novel (novel number-familiar color and shape), (b) the color and shape of elements was novel (familiar number-novel color and shape), and (c) the number, color, and shape of elements was novel (novel number-novel color and shape). For the three types of novel transfer problems, irrelevant dimensions could be both cue-ambiguous, one cue-ambiguous and the other cue-constant, or both cue-constant. The novel numerosities were 3, 5, and 7; thus, six novel numerical identity problems were created ($3 \rightarrow 3$, not 5 or 7; $5 \rightarrow 5$, not 3 or 7; and $7 \rightarrow 7$, not 3 or 5) for each irrelevant dimension type. The novel element shapes were crosses, cylinders, and triangles and the novel element colors were white (W), blue (Bl), and green (G). Figure 5.1.4 shows example problems as a function of the three novel problem types at each of the three irrelevant dimension types.

In completing Phase 2 and finishing Experiment 6, the subject completed 22 sessions such that 1,408 reinforced baseline, 396 nonreinforced baseline, and 396 nonreinforced probe trials (108 novel number-familiar color and shape, 132 familiar number-novel color and shape, and 156 novel number-novel color and shape problems) were completed.

	Novel Number- Novel Color and Shape	Novel Number- Familiar Color and Shape	Familiar Number- Novel Color and Shape
Shape and Color Cue-Ambiguous			
Color Cue-Ambiguous and Shape Cue-Constant			

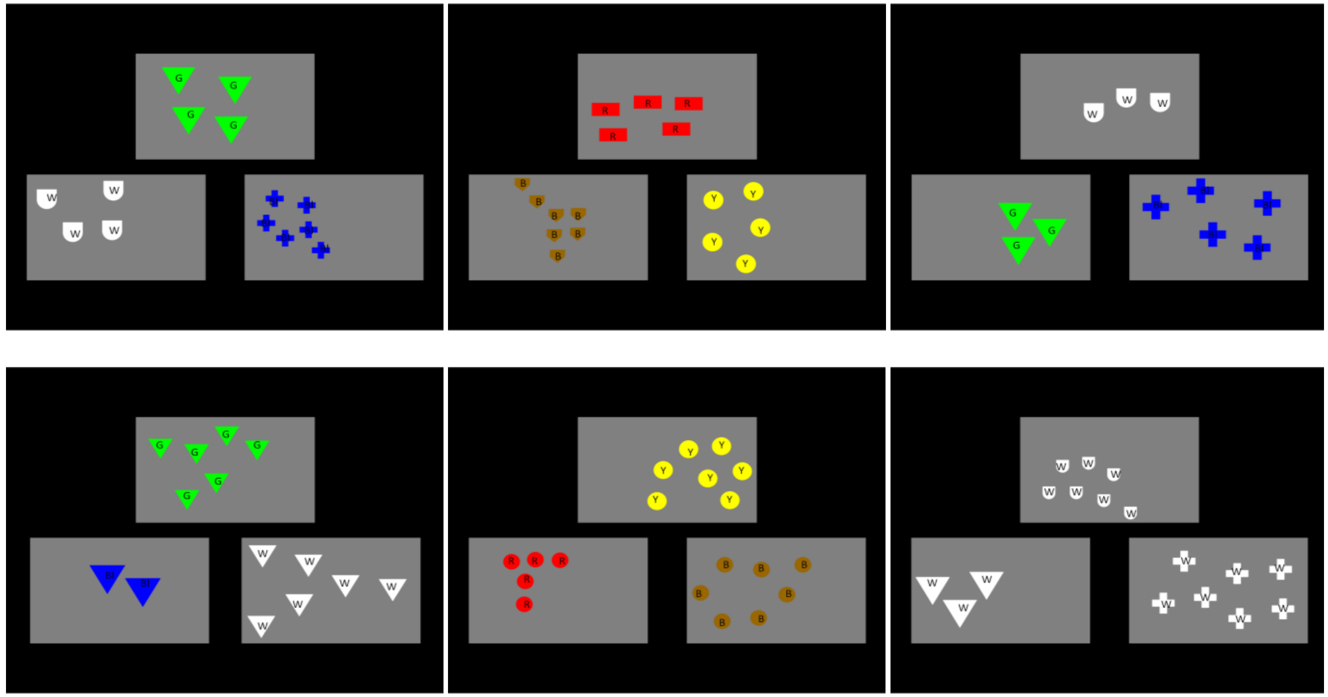


Figure 5.1.4. Example numerical identity transfer problems as a function of novel problem type and irrelevant dimension type for Experiment 6 Phase 2.

Shape Cue-Ambiguous

and

Color Cue-Constant

Color and Shape

Cue-Constant

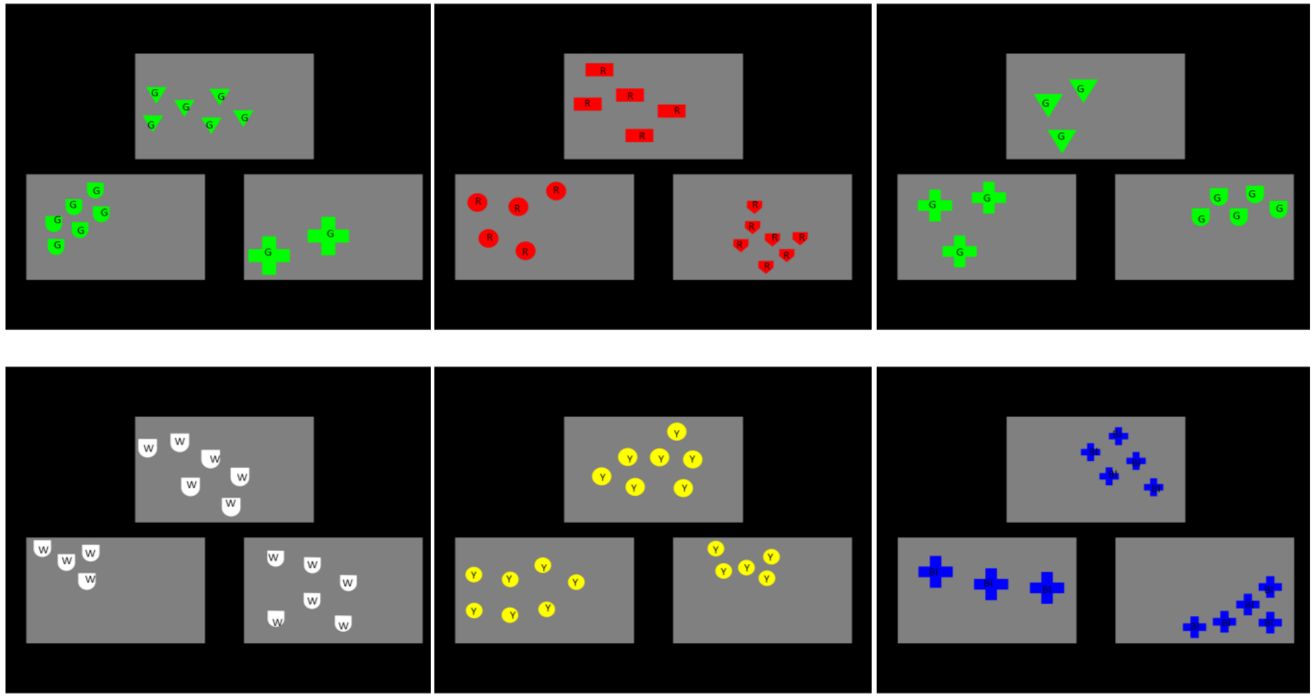


Figure 5.1.4. Continued.

5.2 Results

5.2.1 Data Analysis

For three experiments subdivided into one to three parts (A, B, and C) and/or two to five phases the variables of interest were: (1) subjects (Junior and Madu), (2) problem type (nonidentity and identity), (3) comparison pair type (NOTA absent, as the incorrect comparison, and as the correct comparison), (4) irrelevant dimension type (shape and color both cue-ambiguous, one cue-ambiguous and one cue-constant, and both cue-constant), (5) schedule of reinforcement (CRF and VR-3), and (6) trial type (nonreinforced baseline, reinforced baseline, and nonreinforced probe trials).

The data were analyzed as described in Chapter 3 for Experiments 1, 2, and 3 for the binomial, chi-square, and Fisher's exact tests that were applied to assess the relation between responses (correct vs. incorrect) and the aforementioned variables. Like in the previous experiments, the instances when a session did not contain 100 trials and when subjects received additional sessions after reaching the performance criteria are reported in text, tables, or figures. Regardless, the last 300 trials of a phase or experiment were considered criterion learning and the first 100 trials considered early learning.

Linear regression analyses were conducted to predict the percentage of correct responses (accuracy) and response time (reported in seconds from the onset time of the comparison stimuli to a subject's touch) from various predictor variables. Response times falling outside of two standard deviations from the mean were deemed outliers and removed from the data before linear regression analyses were conducted. Each predictor variable was entered into the analysis using a forward selection method that sequentially entered variables into the model, starting with the one with the largest positive or negative correlation with the dependent variable, but only if they satisfied the criterion for entry (*F*-entry, $p < .05$). Two-tailed *t* tests determined whether the unstandardized regression coefficients (*B*) were statistically different from zero. A criterion α level of .05 was used for linear regression analyses.

The predictor variables concerned the numerical relations among the sample stimulus (S), the correct comparison stimulus (C), and the incorrect comparison stimulus (I). The numerical distance between the correct and incorrect comparison was the absolute value of the algebraic difference between their numerosities (abs C-I). The numerical total of the correct and incorrect comparison was the algebraic sum of their numerosities (C+I). The element area disparity ratio (abs [S-C]/[S-I]) and inter-element distance disparity ratio (abs [S-C]/[S-I]) indicated the degree of divergence between the sample and correct comparison's element area or inter-element distance as a function of the divergence between the sample and incorrect comparison's element area or inter-element distance. Specifically, large disparity ratios indicated that the sample and correct comparison were more different than were the sample and incorrect comparison with respect to element area or inter-element distance. Smaller disparity ratios indicated that the sample and incorrect comparison were more different than were the sample and correct comparison with respect to element area or inter-element distance. *F*-tests assessed model fit in relation to the variables as predictors of the percentage of correct responses and response time with the strength of the relation indexed by R^2.

5.2.1.1 Experiment 4: Concurrent Numerical Identity-Nonidentity Responding with Irrelevant Dimensions Cue-Ambiguous

5.2.1.1.1 Color and Shape Cue-Ambiguous (Phase 1)

Madu did not meet the performance criteria (i.e., above chance [59% correct] accuracy for three consecutive sessions with at least five sessions completed) after completing 4,000 trials (52% correct across all trials). Her accuracy was not different from chance for 35 sessions (51% correct, n = 3,500; binomial tests, ps > .067), below chance for one session (38% correct, n = 100; binomial test, p = .010), and above chance for only four sessions (62% correct, n = 400; binomial tests, ps < .028). Similarly, Junior did not meet the performance criteria after completing 4,000 trials (52% correct across all trials). His accuracy was not different from chance for 36 sessions (50% correct, n = 3,600; binomial tests, ps > .067) and above chance for only four sessions (62% correct, n = 400; binomial tests, ps < .028). The left panel of Figure 5.2.1 displays subject accuracy collapsed across comparison pair type as a function of sessions (as the solid black line).

Chi-square tests revealed a statistically significant and strong relationship between responses and comparison pair type for both Madu and Junior: respectively, χ^2s (2, Ns = 4,000) = 500.37 and 311.89, ps < .001, Vs = .35. For Madu, the pattern was such that her accuracy was significantly higher when the NOTA comparison stimulus was the incorrect comparison (73% correct, n = 1,080) than when NOTA was absent (54% correct, n = 1,840; z = -10.28, p < .001) or when NOTA was the correct comparison (26% correct, n = 1,080; z = 22.16, p < .001); further, her accuracy was significantly higher when the NOTA comparison stimulus was absent than when it was the correct comparison (z = 14.99, p < .001). The opposite pattern characterized Junior's responses.

That is, his accuracy was significantly higher when the NOTA comparison stimulus was the correct comparison (71% correct, n = 1,080) than when NOTA was absent (51% correct, n = 1,840; z = -10.50, p < .001) or was the incorrect comparison (33% correct, n = 1,080; z = -17.66, p < .001); further, his accuracy was significantly higher when the NOTA comparison stimulus was absent than when NOTA was the incorrect comparison (z = 9.52, p < .001). The left panel of Figure 5.2.1 displays subject accuracy for each comparison pair type as a function of sessions.

5.2.1.1.2 Color and Shape Cue-Ambiguous under VR-3 (Phase 2)

Madu did not meet the performance criteria after completing 1,000 trials as her accuracy was at chance for all ten sessions (50% correct, n = 1,000; binomial tests, ps > .097). Likewise, Junior did not meet the performance criteria after completing 1,000 trials (50% correct) as his accuracy did not differ from chance for all ten sessions (binomial tests, ps > .097). The right panel of Figure 5.2.1 displays subject accuracy collapsed across comparison pair type as a function of sessions (as the black solid line).

Chi-square tests revealed a statistically significant and moderately strong relationship between responses and the comparison pair type for both Madu and Junior: respectively, χ^2s (2, Ns = 1,000) = 20.85 and 25.76, ps < .001, Vs = .14 and .16. The pattern was reversed from that found under the CRF schedule in Phase 1 for both Madu and Junior. Specifically, Madu's accuracy was significantly higher when the NOTA comparison stimulus was the correct comparison (60% correct, n = 270) than when it was absent (50% correct, n = 460; z = -2.78, p = .003) or the incorrect comparison (40% correct, n = 270; z = -4.56, p < .001); further, her accuracy was significantly higher when the NOTA comparison stimulus was absent than when it was the incorrect comparison stimulus (z = 2.41, p = .008). Second, Junior's accuracy was significantly higher when the NOTA comparison stimulus was the incorrect comparison (58% correct, n = 270; z = 4.58, p < .001) or when it was absent (53% correct, n = 460; z = 4.03, p < .001) than when the NOTA comparison stimulus was the correct comparison (37% correct, n = 270); however, there was no difference in accuracy between when the NOTA comparison stimulus was incorrect and when the NOTA comparison stimulus it was absent (z = -1.40, p = .082). The right panel of Figure 5.2.1 displays subject accuracy for each comparison pair type as a function of sessions.

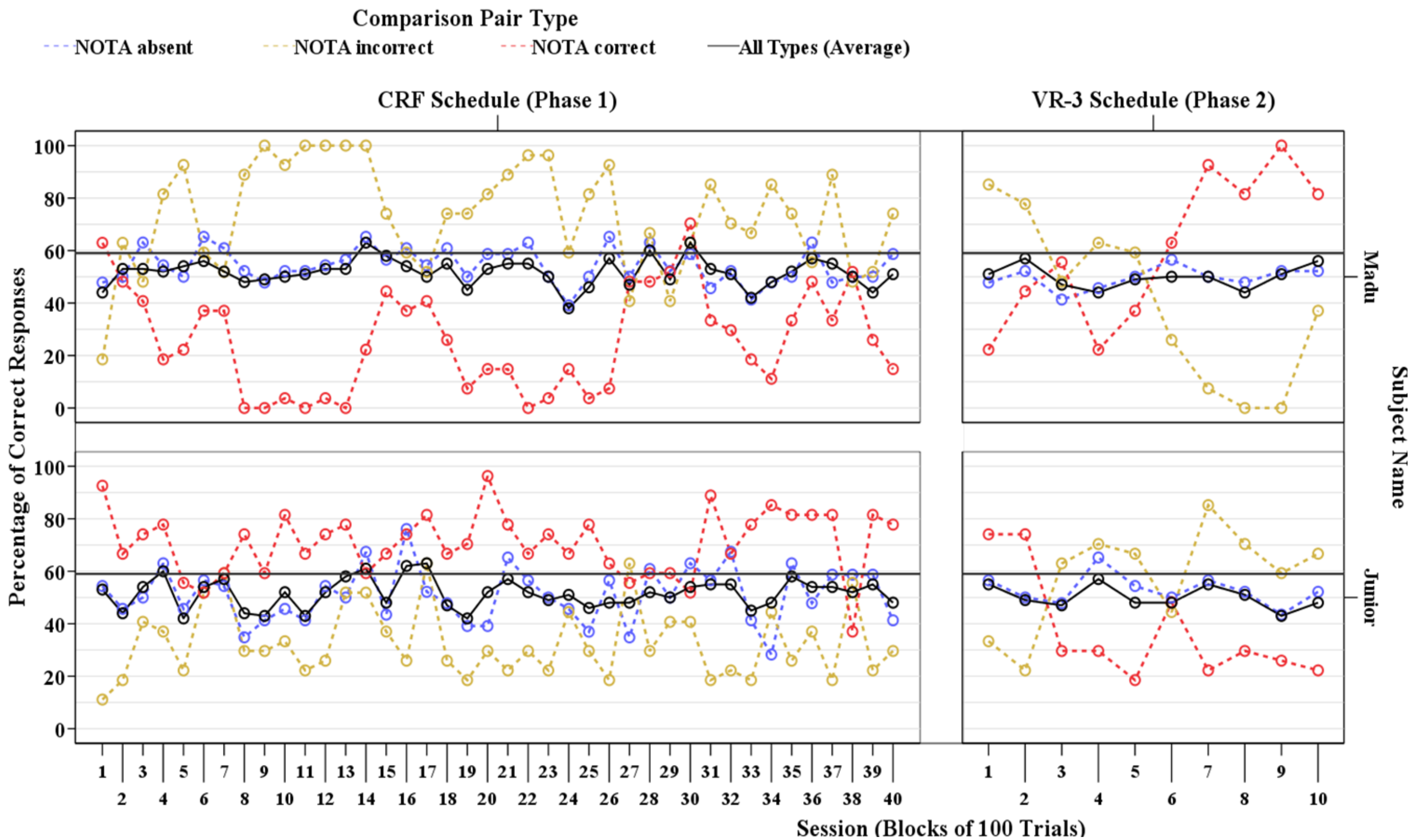


Figure 5.2.1. Subject accuracy as a function of comparison pair type and sessions when both irrelevant dimensions were cue-ambiguous for each phase of Experiment 4. The black horizontal line at 59% correct depicts the lowest percentage of correct responses that was statistically above chance for all comparison pair types (binomial test, $p < .05$, $n = 100$). Note that each session contained 46, 27, and 27 trials of the NOTA absent, NOTA correct, and NOTA incorrect comparison pair types, respectively.

5.2.1.2 Experiment 5A: Numerical Identity Responding with Irrelevant Dimensions Cue-Ambiguous

5.2.1.2.1 Color and Shape Cue-Ambiguous under VR-3 (Phase 1)

Madu did not meet the performance criteria after completing 2,500 trials (51% correct). Her accuracy was not different from chance for 23 sessions (51% correct, n = 2,300; binomial tests, ps > .067), below chance for one session (38% correct, n = 100; binomial test, p = .010), and above chance for only one session (65% correct, n = 100; binomial test, p = .002). Similarly, Junior did not meet the performance criteria after completing 2,500 trials (50% correct). His accuracy was not different from chance for 21 sessions (49% correct, n = 2,100; binomial tests, ps > .097), below chance for two sessions (41% correct, n = 200; binomial tests, ps < .044), and above chance for only two sessions (61% correct, n = 200; binomial tests, ps < .028). The left panel of Figure 5.2.2 displays subject accuracy as a function of sessions.

5.2.1.2.2 Color and Shape Cue-Ambiguous (Phase 2)

Madu did not meet the performance criteria after completing 2,000 trials (49% correct). Her accuracy was not different from chance for 18 sessions (48% correct, n = 1,800; binomial tests, ps > .067), below chance for one session (39% correct, n = 100; binomial test, p = .018), and above chance for only one session (61% correct, n = 100; binomial test, p = .018). Similarly, Junior did not meet the performance criteria after completing 2,000 trials (48% correct). His accuracy was not different from chance for 15 sessions (49% correct, n = 1500; binomial tests, ps > .184), below chance for four sessions (37% correct, n = 400; binomial tests, ps < .044), and above chance for only one session (68% correct, $n = 100$; binomial test, $p < .001$). The second left panel of Figure 5.2.2 displays subject accuracy as a function of sessions.

5.2.1.2.3 Color and Shape Cue-Ambiguous for Sample Numerosity of 4 (Phase 3)

Madu did not meet the performance criteria after completing 500 trials (51% correct). Her accuracy was not different from chance for all five sessions (binomial tests, $ps > .136$). Likewise, Junior did not meet the performance criteria after completing 500 trials (51% correct). His accuracy was not different from chance for four sessions (53% correct, $n = 400$; binomial tests, $ps > .067$) and was below chance for one session (41% correct, $n = 100$; binomial test, $p = .044$). The center panel of Figure 5.2.2 displays subject accuracy as a function of sessions.

5.2.1.2.4 Color and Shape Cue-Ambiguous for Sample Numerosity of 2 (Phase 4)

Madu did not meet the performance criteria after completing 500 trials (54% correct). Her accuracy was not different from chance for all five sessions (binomial tests, $ps > .097$). Similarly, Junior did not meet the performance criteria after completing 500 trials (48% correct). His accuracy also did not differ from chance for all five sessions (binomial tests, $ps > .184$). The second panel from the right of Figure 5.2.2 displays subject accuracy as a function of sessions.

5.2.1.2.5 Color and Shape Cue-Ambiguous for Sample Numerosity of 6 (Phase 5)

Madu did not meet the performance criteria after completing 500 trials (46% correct). Her accuracy was not different from chance for four sessions (48% correct, $n = 400$; binomial tests, $ps > .242$) and below chance for one session (40% correct, $n = 100$; binomial test, $p = .028$). Similarly, Junior did not meet the performance criteria after completing 500 trials (54% correct). His accuracy was not different from chance for four sessions (51% correct, $n = 400$; binomial tests, ps > .136) and was above chance for only one session (64% correct, $n = 100$; binomial test, $p = .003$). The right panel of Figure 5.2.2 displays subject accuracy as a function of sessions.

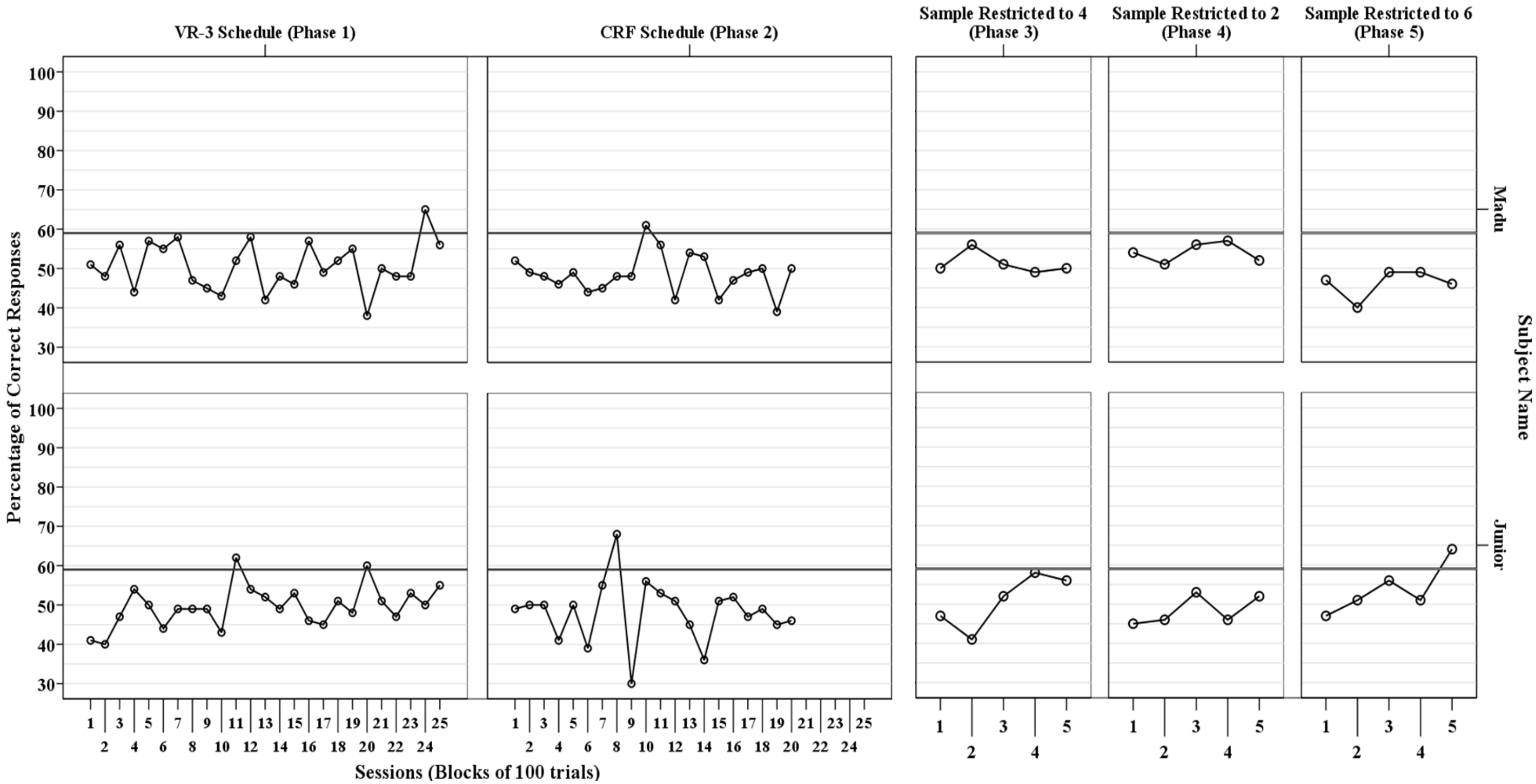


Figure 5.2.2. Subject accuracy as a function of sessions when both irrelevant dimensions were cue-ambiguous for each phase of Experiment 5A. The black horizontal line at 59% correct shows the lowest percentage of correct responses that was statistically above chance (binomial test, $p < .05$, $n = 100$).

5.2.1.3 Experiment 5B: Numerical Identity Responding with Irrelevant Dimensions Cue-Ambiguous and Cue-Constant

5.2.1.3.1 Shape Cue-Ambiguous and Color Cue-Constant (Phase 1)

Madu did not meet the performance criteria after completing 500 trials (54% correct). Her accuracy was not different from chance for four sessions (53% correct, n = 400; binomial tests, ps > .184) and above chance for only one session (59% correct, n = 100; binomial test, p = .044). Similarly, Junior did not meet the performance criteria after completing 500 trials (39% correct). His accuracy was not different from chance for three sessions (45% correct, n = 300; binomial tests, ps > .097) and below chance for two sessions (30% correct, n = 200; binomial tests, ps < .001). The left panel of Figure 5.2.3 displays subject accuracy as a function of sessions.

5.2.1.3.2 Color Cue-Ambiguous and Shape Cue-Constant (Phase 2)

Madu did not meet the performance criteria after completing 500 trials (55% correct). Her accuracy was not different from chance for all five sessions (binomial tests, ps > .067). Likewise, Junior did not meet the performance criteria after completing 500 trials (43% correct). His accuracy did not differ from chance for all five sessions (binomial tests, ps > .309). The right panel of Figure 5.2.3 displays subject accuracy as a function of sessions.

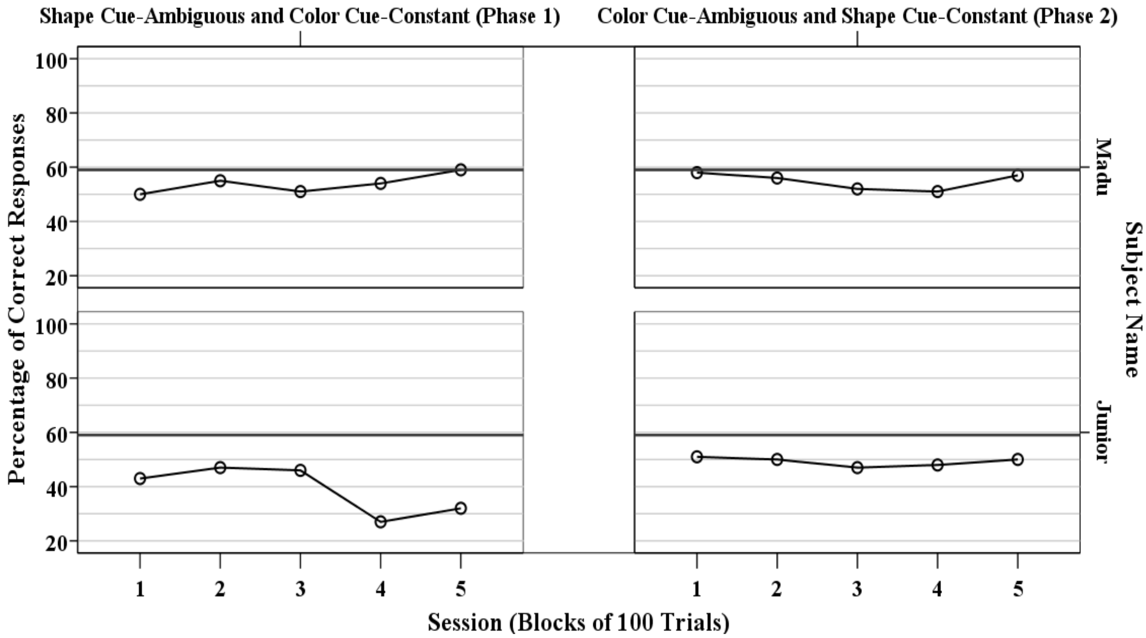


Figure 5.2.3. Subject accuracy as a function of sessions for numerical identity responding when one irrelevant dimension was cue-constant and the other cue-ambiguous for the phases of Experiment 5B. The black horizontal line at 59% correct depicts the lowest percentage of correct responses that was statistically above chance (binomial test, $p < .05$, $n = 100$).

<u>5.2.1.4 Experiment 5C: Numerical Identity Responding with Irrelevant Dimensions Cue-Constant</u>

5.2.1.4.1 Color and Shape Cue-Constant (Phase 1)

Junior did not meet the performance criteria after completing 914 trials (50% correct).[27] His accuracy was not different from chance for eight sessions (50% correct, $n = 714$; binomial tests, *p*s > .184), below chance for one session chance (40% correct, $n = 100$; binomial test, $p = .028$), and above chance for only one session (60% correct, $n =$

[27] During the last session, Junior completed only 14 trials.

100; binomial test, $p = .028$). Junior's participation in Phase 1 concluded his involvement with all subsequent numerical identity tasks. On the other hand, Madu meet the performance criteria (i.e., 59% correct or higher for three consecutive sessions for a minimum of five sessions) after completing 900 trials.[28] She was 60% correct across all ten sessions and her accuracy was above chance for the last four consecutive sessions (62% correct, $n = 400$; binomial tests, ps < .028), above chance for one other session (60% correct, $n = 100$; binomial test, $p = .028$), and not different from chance for five sessions (56% correct, $n = 100$; binomial tests, ps > .067). The left panel of Figure 5.2.4 displays subject accuracy as a function of sessions.

5.2.1.4.2 Color and Shape Cue-Constant for Sample Numerosity of 4 (Phase 2)

Madu met the performance criteria after completing 500 trials (61% correct). Her accuracy was above chance for the last three consecutive sessions (65% correct, $n = 300$; binomial tests, ps < .010) and at chance for two sessions (55% correct, $n = 200$; binomial tests, *ps* >.097). The second panel from the left of Figure 5.2.4 displays Madu's accuracy as a function of sessions.

5.2.1.4.3 Color and Shape Cue-Constant for Sample Numerosity of 6 (Phase 3)

Madu met the performance criteria after completing 500 trials (79% correct). Her accuracy was above chance for all five sessions (binomial tests, ps < .001). The second center panel of Figure 5.2.4 displays Madu's accuracy as a function of sessions.

5.2.1.4.4 Color and Shape Cue-Constant for Sample Numerosity of 2 (Phase 4)

---

[28] Madu received an additional 100-trial session after meeting the performance criteria.

Madu met the performance criteria after completing 500 trials (80% correct) as her accuracy was above chance for all five sessions (binomial tests, *p*s < .001). The second panel from the right of Figure 5.2.4 displays Madu's accuracy as a function of sessions.

5.2.1.4.5 Color and Shape Cue-Constant (Phase 5)

Madu met the performance criteria after completing 500 trials (69% correct) and her accuracy was above chance for all five sessions (binomial tests, *p*s < .010). Further, the subject was 73% correct across the 1,500 additional trials that she completed after reaching the performance criteria, with her accuracy above chance for each of the 15 sessions (binomial tests, *p*s < .002). The right panel of Figure 5.2.4 displays Madu's accuracy across sessions. After completing Phase 5, Madu finished Experiment 5.

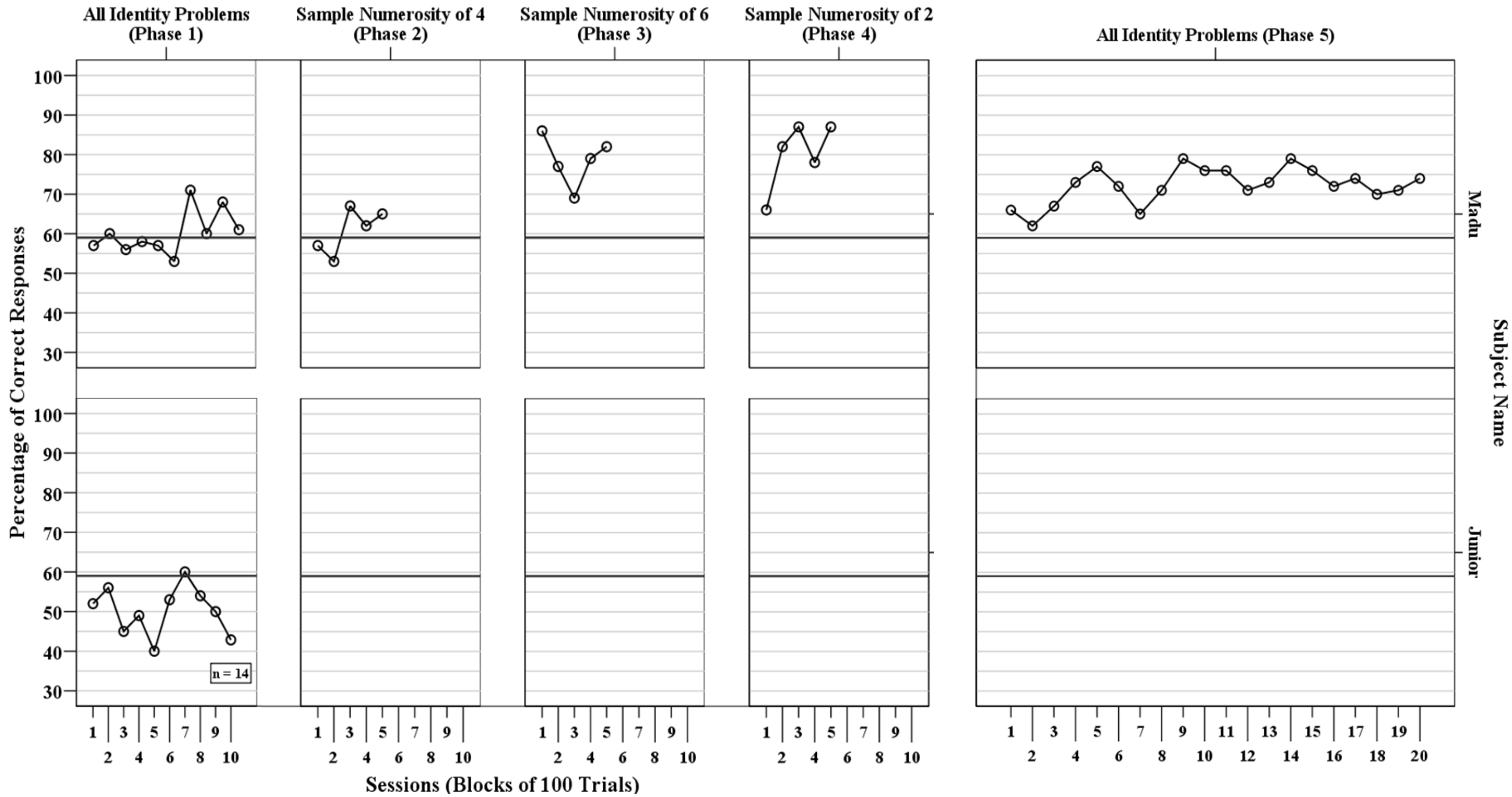


Figure 5.2.4. Subject accuracy as a function of sessions during numerical identity responding when both irrelevant dimensions were cue-constant for the phases of Experiment 5C. The black horizontal line at 59% correct depicts the lowest percentage of correct responses that was statistically above chance (binomial test, $p < .05$, $n = 100$).

5.2.1.4.6 Early and Criterion Learning Among Experiment Phases

To examine Madu's performance between consecutive phases in relation to early and criterion learning, the proportion of correct responses during the final three sessions of Phases 1, 2, 3, and 4 were compared to the first session of the phase that followed it (i.e., respectively Phases 2, 3, 4, and 5). A chi-square test revealed a statistically significant, moderately strong relationship between responses and sessions, χ^2 (7, N = 1,600) = 67.65, $p < .001$, $V = .21$. The pattern was such that there was no difference between criterion learning in Phase 1 and early learning in Phase 2 (63% vs. 57% correct, n = 300 and 100; $z = 1.07$, $p = .143$), but accuracy was significantly lower during criterion learning in Phase 2 than early learning in Phase 3 (65% vs. 86% correct, n = 300 and 100; $z = -4.03$, $p < .001$), significantly higher during criterion learning in Phase 3 than early learning in Phase 4 (77% vs. 66% correct, n = 300 and 100; $z = 2.11$, $p = .018$), and significantly higher during criterion learning in Phase 4 than early learning in Phase 5 (84% vs. 66% correct, n = 300 and 100; $z = 3.86$, $p < .001$).

Additionally, the final three sessions of Phases 1, 2, 3, 4, and 5 were used to assess Madu's criterion level performance in relation to the six numerical identity problems. A chi-square test indicated that there was a statistically significant, moderately strong relationship between responses and the numerical identity problems, χ^2 (5, N = 1,500) = 117.45, $p < .001$, $V = .28$. Accuracy was highest for (zs > 2.72, ps < .003), but not different between ($z = 0.96$, $p = .168$) the following two problems: 6 → 6, not 2 and 2 → 2, not 6. Accuracy was lowest for 4 → 4, not 6 (zs > 3.61, ps < .001): Finally, there were no differences between 4 → 4, not 2 and 2→ 2, not 4 ($z = -0.62$, ps = .269) and between 4 → 4, not 2 and 6 → 6, not 4 ($z = 2.12$, ps = .017 adjusted). Figure 5.2.5

displays accuracy for Madu as a function of the six numerical identity problems during criterion learning for Experiment 5C.

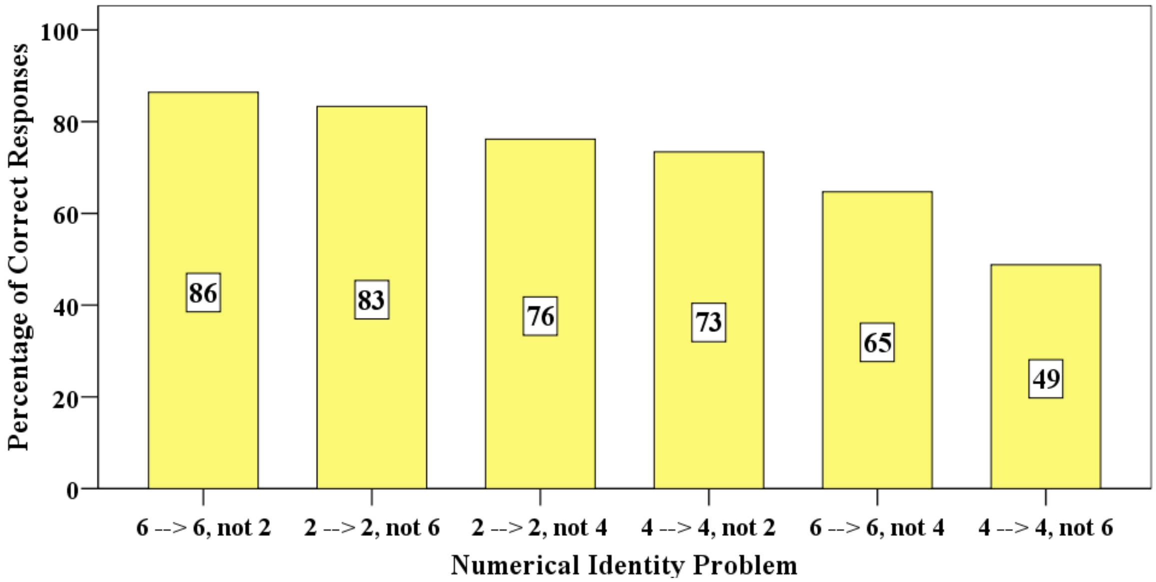


Figure 5.2.5. Accuracy as a function of the numerical identity problems for Madu during criterion learning when both irrelevant dimensions were cue-constant for the phases of Experiment 5C.

Linear regression analyses were conducted to assess linear trends in Madu's accuracy and response times during criterion learning for the phases of Experiment 5C. The data from Experiment 5C during the final three sessions of Phases 1, 2, 3, 4, and 5 were included in the analyses. The nine trials that Madu received during criterion learning in which element area was the same for the correct and incorrect comparison stimulus were removed from these analyses. The percentage of correct responses and average response time were calculated for every possible combination of numerosities, element areas, and inter-element distances among the sample, incorrect, and correct comparison stimuli (N = 226).

Both numerical distance and numerical total of the comparison stimuli were significantly related to the percentage of correct responses in the final regression model, $F(2, 223) = 33.91$, $p < .001$. The element area and inter-element distance disparity ratios did not meet the F-entry criterion; thus, they were not entered into the regression model. The final regression model explained a moderate amount of variance ($R^2 = .23$). T tests indicated that the slope was significantly different from zero for the numerical distance between comparisons ($B = 9.36$, $SE = 1.35$; $t = 6.94$, $p < .001$) and the numerical total of comparisons ($B = -3.84$, $SE = .80$; $t = -4.80$, $p < .001$). The relationships were such that accuracy was higher when the numerical difference between comparisons was largest (difference of 4: 84% correct, $SE = 2.00$) rather than smallest (difference of 2: 66% correct, $SE = 1.77$) and accuracy increased as the numerical total of comparisons decreased from a numerosity of 10 (58% correct, $SE = 2.55$), to a numerosity of 8 (84% correct, $SE = 2.00$), to a numerosity of 6 (73% correct, $SE = 2.17$).

For the second regression conducted, the numerical distance and element area disparity ratio between comparisons was significantly related to response time in the final regression model, $F(2, 225) = 9.35$, $p < .001$. The numerical total and inter-element distance disparity ratio did not meet the F-entry criterion; thus, they were not entered into the regression model. The final regression model explained only a small amount of variance ($R^2 = .08$). T tests indicated that the slope was significantly different from zero for numerical distance ($B = -.05$, $SE = .02$; $t = -3.42$, $p = .001$) such that response times were longer when the numerical difference between comparisons was smallest (difference of 2: 1.27 s, *SE* = .02) rather than largest (difference of 4: 1.16 s, *SE* = .02) and that the slope was significantly different from zero for the element area disparity ratio (B = .03, SE = .02; t = 2.16, p = .032) such that response times increased as the element area disparity ratio increased.

5.2.1.5 Experiment 6: Numerical Identity Responding and Transfer Test

5.2.1.5.1 Reinforced and Nonreinforced (36%) Baseline (Phase 1)

Madu met the performance criteria after completing 500 trials. Her accuracy was above chance for all five sessions (71% correct; binomial tests, ps < .001). For the 1,500 additional trials she completed after she reached the criterion, the subject's accuracy also was above chance for each session (74% correct; binomial tests, ps < .001). With respect to all 20 sessions, she was 73% correct for reinforced (n = 1,520) and 74% correct for nonreinforced baseline trials (n = 480). In particular, her accuracy was above chance with the 64 reinforced baseline trials in each of the 20 sessions (binomial tests, ps < .017) and above chance with the 36 nonreinforced baseline trials in 18 of the 20 sessions (binomial tests, ps < .033). The left panel of Figure 5.2.6 displays Madu's accuracy as a function of sessions for each trial type.

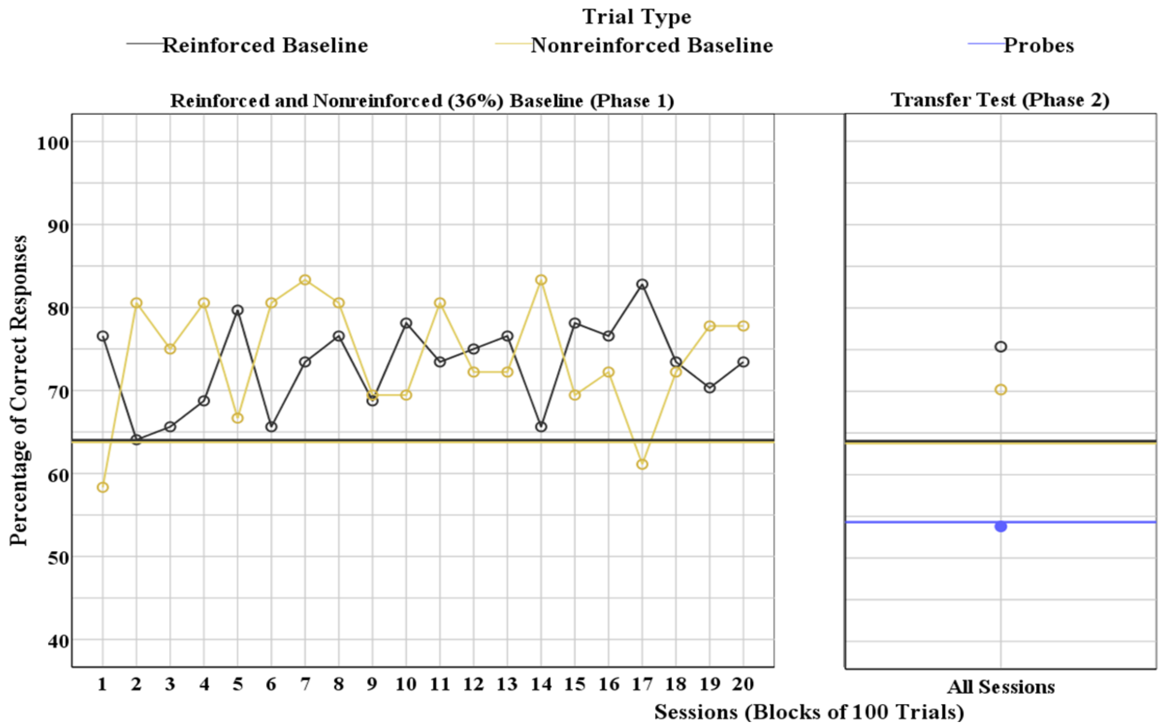


Figure 5.2.6. Accuracy for Madu as a function of sessions and trial type when irrelevant dimensions were cue-constant during the phases of Experiment 6. The black and gold horizontal line at 59% correct depicts the lowest percentage of correct response that was statistically above chance for the average accuracy of reinforced and nonreinforced baseline trials in Phase 1 (binomial test, $p < .05$, $n = 100$); likewise with the blue line at 54% correct for nonreinforced probe trials in Phase 2 (binomial test, $p < .05$, $n = 396$).

There was a statistically significant, moderately strong relationship between responses and the numerical identity problems during the last 300 trials of reinforced and nonreinforced baseline trials, χ^2 (5, $N = 300$) = 144.26, $p < .001$, $V = .27$. Accuracy was highest for 6 → 6, not 2 (zs > 3.34, ps < .001), but not different from 2 → 2, not 6 (z = 2.12, p = .017 adjusted). Accuracy was lowest for 4 → 4, not 6 (zs > 2.65, ps < .004), but not different from 6 → 6, not 4 (z = 2.01, p = .022 adjusted) or 2 → 2, not 4 (z = 1.78, p =

.038 adjusted). No other pairwise comparison was significant. Figure 5.2.7 displays accuracy for Madu as a function of the six numerical identity problems.

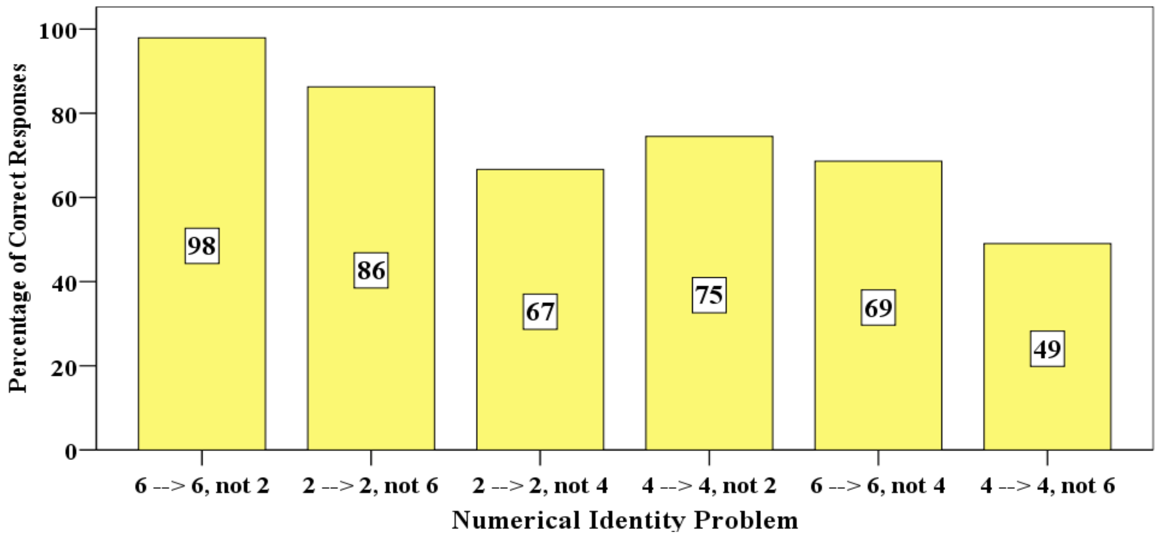


Figure 5.2.7. Accuracy as a function of the six numerical identity problems for Madu during criterion learning when both irrelevant dimensions were cue-constant for Phase 1 of Experiment 6.

Two regressions were conducted to assess linear trends for numerical identity matching accuracy and response time across all twenty sessions. The percentage of correct responses and average response times were calculated for every possible combination of numerosities, element areas, and inter-element distances among the sample, incorrect, and correct comparison stimuli (*N* = 199).[29]

---

[29] Criterion performance (i.e., the last 300 trials) was not used to prevent instances where a single trial represented the percentage of correct responses or average response time.

Both the numerical distance and the numerical total between comparisons were significantly related to the percentage of correct responses in the final regression model, $F(2, 196) = 47.09$, $p < .001$. The element area disparity ratio and inter-element distance disparity ratio did not meet the *F*-entry criterion; thus, they were not entered into the regression model. The final regression model explained a moderate amount of variance ($R^2 = .33$). *T* tests indicated that the slope was significantly different from zero for the numerical distance between comparisons ($B = 10.94$, $SE = 1.29$; $t = 8.46$, $p < .001$) and for the numerical total of comparisons ($B = -3.51$, $SE = .74$; $t = -4.71$, $p < .001$). The relationships were such that accuracy was higher when the numerical difference between comparison sets was largest (difference of 4: 88% correct, $SE = 1.20$) rather than smallest (difference of 2: 66% correct, $SE = 1.81$) and accuracy increased as the numerical total of comparisons decreased from a numerosity of 10 (59% correct, $SE = 2.49$) to a numerosity of 8 (88% correct, $SE = 1.20$) to a numerosity of 6 (74% correct, $SE = 2.37$).

For the second regression, only the element area disparity ratio was significantly related to response time in the final regression model, $F(1, 197) = 4.16$, $p = .043$. The numerical distance, numerical total, and inter-element distance disparity ratio did not meet the *F*-entry criterion. The final regression model explained only a small amount of variance ($R^2 = .02$). *T* tests indicated that the slope was significantly different from zero ($B = .02$, $SE = .01$; $t = 2.04$, $p = .043$) such that response times increased as the element area disparity ratio increased.

5.2.1.5.2 Transfer Test (Phase 2)

The right panel of Figure 5.2.6 displays Madu's accuracy during the transfer of learning test. Performance was evaluated across all 22 transfer test sessions for reinforced baseline trials, nonreinforced baseline trials, and nonreinforced probe trials. These analyzes revealed that Madu responses exceeded chance with reinforced baseline trials (75% correct, n = 1,408; binomial test, ps < .001) and nonreinforced baseline trials (70% correct, n = 396; binomial test, p < .001), but were not different from chance for nonreinforced probe trials (54% correct, n = 396; binomial test, p = .073).

Finally, analyzing nonreinforced probe trial accuracy across all 22 transfer sessions as a function of novel problem type and irrelevant dimension type revealed that Madu responded above chance with novel number-familiar cue-constant color and shape problems (67% correct, n = 33; binomial test, p = .040) and novel number-novel cue-constant color and shape problems (65% correct, n = 57; binomial test, p = .017). Further, accuracy for novel number-familiar cue-constant color and shape problems and novel number-novel cue-constant color and shape problems (66% correct, n = 90) did not differ from the subject's criterion level accuracy with reinforced (72% correct, n = 192) and nonreinforced baseline trials (76% correct, n = 108) during the last 300 trials of Phase 1, χ^2 (2, N = 390) = 2.68, p = .263. The subject's accuracy did not differ from chance for any other novel problem-irrelevant dimension type combination (binomial tests, ps > .095). Figure 5.2.8 displays the subject's accuracy during the transfer of learning test for the combinations of novel problem type and irrelevant dimension type.

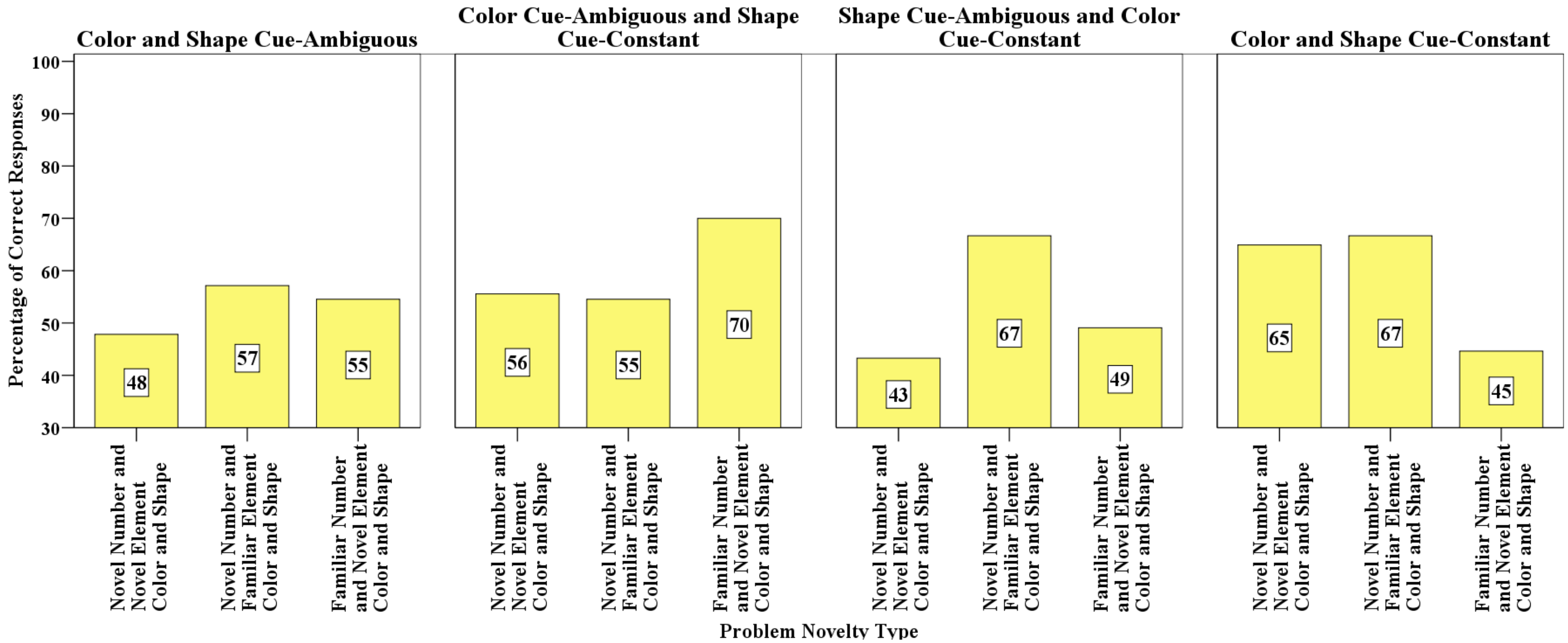


Figure 5.2.8. Accuracy for Madu during the transfer test of Experiment 6 for novel problem type as a function of the irrelevant dimension type.

5.3 Discussion

Neither subject responded correctly to concurrent numerical identity and nonidentity problems after completing 4,000 trials in Experiment 4 when the irrelevant dimensions of color and shape were cue-ambiguous. Instead, they both exhibited avoidance and preference response patterns for the NOTA comparison stimulus and responded randomly when NOTA was not present as a comparison. Under CRF, Madu avoided selecting the NOTA comparison when it was present in the comparison stimulus pair regardless of whether it was correct to do so. The result was that her accuracy was high and above chance when the NOTA comparison was the incorrect stimulus and low and below what chance would predict when the NOTA comparison was the correct stimulus. On the other hand, under CRF, Junior preferred to select the NOTA comparison stimulus when it was present in the comparison stimulus pair regardless of whether it was correct to do so. The result was that his accuracy was high and above chance when the NOTA comparison was the correct stimulus and low and below chance when the NOTA was the incorrect stimulus. Interestingly, when the reinforcement schedule switched from CRF to VR-3, both subjects swapped response patterns; that is, Madu began to prefer to choose the NOTA comparison and Junior began to avoid the NOTA comparison. Finally, aside from preference and avoidance for the NOTA, both subjects responded randomly when only numerosity stimuli were present as comparisons before and after the reinforcement schedule changed from CRF to VR-3.

Explaining the aforementioned pattern of preference and avoidance is difficult. If whole-nonidentity learning about color transferred from the earlier experiment detailed in Chapter 3, then subjects should have preferred to select the NOTA comparison whenever it was present because the irrelevant dimensions were cue-ambiguous for identity and nonidentity problems so the color (and the shape) of a comparisons' elements never matched the sample. This is what Junior did at first under the CRF schedule and what Madu switched to do when the reinforcement schedule changed to VR-3 so transfer of learning about the NOTA comparison stimulus in terms of color whole-identity matching cannot account for the response patterns demonstrated in the present experiment by subjects.

The change in reinforcement schedule seemed to prompt the switch in subject response pattern, which demonstrates that the orangutans were sensitive to the relative rate of reinforcement. Two different reinforcement schedules were utilized during the concurrent numerical identity and nonidentity tasks in Experiment 4 and during the first part of Experiment 5 because the previous work of other authors suggested that the schedule of reinforcement during conditional discrimination tasks affects acquisition. One report demonstrated that three monkeys learned conditional position discrimination problems that involved pairs of colors rapidly under CRF, but their performance was marked by repeated drops to chance accuracy levels; whereas, their performance under a VR schedule illustrated slow, but consistent increases in accuracy (Fujita, 1985). On the other hand, the conditional discrimination accuracy of pigeons was near chance under a CRF schedule, but became higher under an intermittent reinforcement schedule during acquisition (Ferster, 1960).

After receiving 6,000 more trials during Experiment 5A, subjects did not respond above what chance would predict when the irrelevant dimensions remained cue-ambiguous even though the NOTA stimulus was removed from the set of comparisons to restrict the task to numerical identity responding, the problems reduced from all six problems to sets of two problems with the same sample numerosity presented in sequential trial blocks, and the reinforcement schedule changed from VR-3 to CRF. Because reliable discrimination did not emerge under either reinforcement schedule, the present experiment does not provide information about the effectiveness of continuous and partial reinforcement schedules in establishing conditional discrimination in nonhuman primates.

In any case, after receiving 1,000 more trials under CRF during Experiment 5B, subjects still did not respond reliably above what chance would predict when one irrelevant dimension was made cue-constant while the other remained cue-ambiguous. Finally, when both irrelevant dimensions were made cue-constant and CRF still employed during Experiment 5C, Junior ceased to respond after completing about 900 trials with his accuracy remaining around 50% correct during the first phase. Conversely, after completing 900 trials during the first phase of Experiment 5C, Madu reached the performance criteria by obtaining above chance accuracy (59% correct or better) for more than three consecutive sessions. Her accuracy continued to improve across the next 1,500 trials when the set of numerical identity problems was again reduced from all six problems to sets of two problems that had the same sample numerosity that were presented in sequential trial blocks during the second, third, and fourth phases. During the final phase of Experiment 5C, Madu judged numerical identity with all six problems above what chance would predict. She reached an accuracy of 69% correct after completing 500 trials and her accuracy stabilizing at around 73% correct after completing an additional 1,500 trials. In total, thus, Madu completed 15,500 trials while learning to respond to numerical identity for six problems.

Numerical identity responding only occurred when irrelevant dimensions were cue-constant, in other words, when the color and shape of the elements that instantiated a numerosity were identical between all stimuli being compared. That Madu did not learn to match numerical identity or nonidentity when the irrelevant dimensions were not cue-constant suggests that numerical identity responding must first be established with between-stimulus variability at its lowest level before it can be established at higher levels. Confirmation of the aforementioned may be provided in the future if Madu demonstrates an ability to respond to numerical identity for the same six problems or subsets of the six problems when at least one irrelevant dimension is cue-ambiguous. The establishment of numerical identity responding with between-stimulus variability at its highest level before it is established at the lowest level has not yet been demonstrated in nonhuman primates (in Chapter 4, c.f., Cantlon & Brannon, 2007; Merritt, Rugani, & Brannon, 2009; Woodruff & Premack, 1981).

During criterion learning when both color and shape were cue-constant during the five phases of Experiment 5C, Madu was least accurate when judging numerical identity for the 4 → 4, not 6 problem, which is defined by the largest numerical size and the smallest numerical distance. Indeed, both the numerical distance and the numerical size of comparisons were significant predictors of accuracy such that accuracy was highest when the difference in the number of element between comparisons was large and when the total number of the elements for comparisons was small. Additionally, both numerical distance and the element area disparity ratio predicted the subject's latency to respond, although they were only weakly predicative. The effects were such that Madu took longer to respond when the numerical distance between stimuli was small rather than large and when the element area disparity ratio was larger rather than smaller. Large element area disparity ratios indicate that the sample and correct comparison's element areas were more different than were the sample and incorrect comparison's element areas. Madu went on to obtain above chance accuracy when 36% of the familiar, trained numerical identity problems were not followed by reinforcement during the first phase of Experiment 6. Again, the subject's numerical identity matching ability was inferior for 4 → 4, not 6, numerical distance and the numerical size were significant predictors of accuracy, and the element area disparity ratio was a significant but weak predictor of response latency. It is important to note that the element area disparity ratio and inter-element distance disparity ratio were never significant predictors of criterion accuracy during the five phases of Experiment 5C or when 36% of the familiar, trained numerical identity problems were not followed by reinforcement during the first phase of Experiment 6. These continuous quantities were experimentally controlled to prevent them from influencing accuracy even though it seems that one exerted an effect on response latency.

The aforementioned numerical magnitude and distance effects coupled with Madu's ability to respond to numerical identity with problems that included 4 and 6 elements, is consonant with the idea that cardinality is not encoded exactly as object-files or otherwise, but instead is encoded as magnitudes with scalar variability as proposed in the accumulator model (Gallistel & Gelman, 2000). Similarly, the accuracy of one language-trained chimpanzee called Ai showed the same kind of numerical distance and numerical size effects (Matsuzawa, 1985; Matsuzawa et al., 1986) and the accuracy and response latency of four monkeys showed the same kind of numerical distance effects (Merritt, Rugani, & Brannon, 2009) during their training on numerical identity tasks.

The final topic for discussion is concept formation. During the final phase of Experiment 6, numerical identity responding partially transferred to 2 of the 12 transfer test novel problem-irrelevant dimension types despite the small number of training numerosities, colors, and shapes employed. The present study utilized six numerical identity problems that employed three different numerosities, three different colors, and three different shapes during training to establish the conditional discrimination; thus, the universe of numerosities, colors, and shapes was not narrowed substantially. This means that the transfer of learning that occurred is not likely because the novel transfer stimuli physically resembled one or more of the training stimuli.

Concept formation occurred for 2 of the 3 transfer test problems that were defined by cue-constant irrelevant dimensions; that is, for novel number-familiar colors and shapes and for novel number-novel colors and shapes, but not for familiar number-novel colors and shapes. Acquisition of the conditional discrimination occurred only when the irrelevant dimensions of problems were both cue-constant; thus, it is likely that the generality of concept formation was restricted to the same domain such that concepts did not form for problems that had one or more cue-ambiguous irrelevant dimensions. Restricted-domain relational learning is defined as being able to perform a task relationally within limits that are circumscribed by the training stimuli and increasing the training set of stimuli is the mechanism that expands the domain (Wright & Katz, 2009; Wright & Lickteig, 2010).

In the introduction to the dissertation, I asserted that comparative examinations between all sister ape species and nonverbal human infants are necessary to discover information about the origins, mechanisms, and development of primate cognition. Operant procedures have not yet been utilized with nonverbal human infants to discover how they learn about and conceptualize numerical identity; furthermore, the research reports that employ habituation/dishabituation and novelty procedures to investigate numerical cognition do not document the language production and comprehension abilities of their infant participants. Nonhuman primates and nonverbal human infants may share the same set of basic numerical competencies, in particular, the ability to relationally learn and form domain-specific concepts about numerical identity.

That Madu continued to match based on numerical identity for the majority of novel numerical identity problems that fell within the domain of her earlier learning illustrates an ability to form a conceptual understanding of cardinal number identity for the first time in a nonlanguage-trained ape without the use of symbols and between sets of numerosities as indexed by successful transfer of learning to novel numerosities instantiated in familiar and novel element colors and shape. As such, this experiment provides evidence that converges with what was previously demonstrated in an enculturated chimpanzee with symbols and in language-trained chimpanzees with and without the use of symbols (Boysen & Berntson, 1989; Matsuzawa, 1985; Matsuzawa et al., 1986; Woodruff & Premack, 1981) and in monkeys (Cantlon & Brannon, 2007; Merritt, Rugani, & Brannon, 2009).

CHAPTER 6

CONCLUSION

It has been said that detecting identity and nonidentity is central to human cognition (Wasserman & Young, 2010), but few researchers attempt to account for exactly how organisms learn relational concepts about identity and nonidentity from their experiences. It is a hard proposition because any specific example of a relation is always instantiated with some specific set of arguments (e.g., a specific object is above another specific object) so it is never possible to observe an example of a pure disembodied relation (Mandler, 2000). A handful of researchers have originated psychological and neural network models to answer questions about relational learning and concept formation.

One of the recent models to account for how relational concepts form from specific examples was proposed by Doumas, Hummel, and Sandhofer (2008). According to their discovery of relations by analogy (DORA) model, during comparison the properties that objects share become more active than the properties unique to one object or the other. New tokens for the shared roles or objects are connected to the most active features to reflect the feature overlap explicitly. For example, when a child thinks about an elephant and a truck simultaneously, it activates their constituent features (elephants are big, gray, and have trunks, whereas, trucks are big, metallic, and have wheels). The feature that is shared by the elephant and truck (i.e., big) receives twice as much input and become twice as more active than unshared features and then, a new unit learns the connection to the most active feature and links it together. Applied iteratively, DORA results in progressively more refinements and eventually in multi-place relational structures.

This model highlights the importance of part-identity in the discovery of relations. Indeed, our world is full of objects that differ from one another across multiple stimulus dimensions (Evans & Smith, 1988; Lea & Wills, 2008; Smith, 1989). Developmentally, relational learning begins with the discovery of global identity in early childhood, but later the ability extends to include discovering identity between shared common attributes (Burns, 1992; Kemler, 1983; Smith, 1984; Smith, 1993; Smith & Heise, 1992). Even so, it does not follow that humans and nonhuman animals will automatically use all the dimensional information available to them to categorize or conceptualize things (Lea & Wills, 2008). The series of experiments within this dissertation showed that both orangutans learned to judge color whole-identity and nonidentity. They went on to learn to judge color and shape whole- and part-identity, but the extent to which they did so differed in terms of the type of identity (part vs. whole) and the relevant dimension (color vs. shape), and they were even able to conceptualize some identity relations.

Our world is also full of collections of objects that differ from one another in their cardinality. Again, it does not follow that humans and nonhuman animals automatically will use cardinal number to categorize or conceptualize collections of things. The series of experiments within this dissertation showed that one orangutan failed to judge numerical identity at first when between-stimulus variability was at the highest and intermediate level (i.e., four brown circles is the same as four blue triangles), but succeeded in doing so when between-stimulus variability was at its lowest (e.g., four brown circles is the same as four brown circles). Additionally, this subject was able to conceptualize numerical identity, but only in a domain-specific way.

Finding that conditional discrimination failed to transfer to some novel stimuli, but successfully transferred to other novel stimuli for the two orangutans, supplies some support for restricted-domain relational learning. Restricted-domain relational learning is the idea that relational learning can operate within a restricted portion of the stimulus domain that is defined by the characteristics of the training stimuli (Wright & Katz, 2009; Wright & Lickteig, 2010). It is a different way of thinking about concept learning that somewhat paradoxically concludes relational learning in the absence of transfer of learning to novel stimuli. What this means is that concept formation may not be an all-or-none phenomena, but be best characterized as a domain that can expand.

Number is typically treated as its own separate conceptual process, but perhaps it is better thought of as a constituent property of a collection of things just like color and shape can be constituent properties of objects. Indeed, differentiating numerical information from other kinds of information and differentiating the relations between numerical information from the relations between other kinds of information are components involved in the posited developmental progression of numerical knowledge in human infants. Specifically, numerical information is theorized to exist first in unordered subitized states for which there is no understanding that N-ness concerns information about the same kind of thing. In other words, oneness and twoness are initially unrelated to each other such that oneness is different from twoness just as oneness is different from blueness. The next capability added to the system involves infants realizing that N-ness concerns information about the same kind of thing. In other words, infants learn that numerical information is an independent characteristic of stimuli such that the relation among different N-nesses are relations of the same type that are unlike other kinds of relations (e.g., the relation between blueness and blueness is not the same as the relation between oneness and oneness). This provides the basis for generalizing the concept of equality of number from one small numerosity to another (Cooper, 1984; Wynn, 1992a) (but see also, Brainerd, 1979; Simon, 1997; Strauss & Curtis, 1984).

Are primates predisposed to learn about and conceptualize numerical identity in a way that is different from the way that they learn about and conceptualize other things (Dehaene & Changeux, 1993; Geary, 2000; Wynn, 1992b, 1998a, 1998b)? This paper does not claim to provide an answer to this question, but the patterns in how the orangutans responded to and conceptualized identity support the notion that recognizing identity is more difficult when the stimuli being compared are more divergent, regardless of whether the stimuli are numerical or not. So it is harder to judge part-identity and it is harder to judge numerical identity when irrelevant dimensions are cue-ambiguous because the compared sets are defined by a high level of between-stimulus variability. This dissertation found that two orangutans were able to relationally learn about and conceptualize identity and sometimes nonidentity in terms of color, shape, and cardinal number with some constraints. To generate a complete picture about the mechanisms, origins, and development of cognition, it is essential to continue to investigate these types of conceptual abilities in all of the sister ape species.

REFERENCES

Astley, S. L. & Wasserman, E. A. (1996). Mediating associations, essentialism, and non-similarity-based categorization. In T. R. Zentall & P. M. Smeets (Eds.), *Stimulus class formation in humans and animals* (pp. 111-133). New York, NY: Elsevier Science Publishers.

Barros, R. D. S., Galvão, O. d. F., & McIlvane, W. J. (2002). Generalized identity matching-to-sample in *Cebus apella*. *The Psychological Record, 52*(4), 441.

Bell, E. T. (1937). *Men of mathematics*. New York, NY: Simon and Schuster.

Benedict, H. (1979). Early lexical development: Comprehension and production. *Journal of Child Language, 6*(2), 183-200.

Beran, M. J. (2010). Use of exclusion by a chimpanzee (*Pan troglodytes*) during speech perception and auditory–visual matching-to-sample. *Behavioural Processes, 83*(3), 287-291.

Beran, M. J., Evans, T. A., & Harris, E. H. (2008). Perception of food amounts by chimpanzees based on the number, size, contour length and visibility of items. *Animal Behaviour, 75*(5), 1793-1802.

Beran, M. J. & Washburn, D. A. (2002). Chimpanzee responding during matching to sample: Control by exclusion. *Journal of the Experimental Analysis of Behavior, 78*(3), 497-508.

Berger, K. S. (2008). *The developing person through the life span* (7th ed.). New York: Worth Publishers.

Bernstein, I. S. (1961). The utilization of visual cues in dimension-abstracted oddity by primates. *Journal of Comparative and Physiological Psychology, 54*(3), 243-247.

Bhatt, R. S., Wilk, A., Hill, D., & Rovee-Collier, C. (2004). Correlated attributes and categorization in the first half-year of life. *Developmental Psychobiology, 44*(2), 103-115.

Bornstein, M. H. (1984). A description taxonomy of psychological categories used by infants. In C. Sophian (Ed.), *Origins of cognitive skills* (pp. 313-338). Hillsdale, NJ: Lawrence Erlbaum Associates.

Boysen, S. T. (1997). Representation of quantities by apes. In P. Slater, C. Snowdon, J. Rosenblatt & M. Milinski (Eds.), *Advances in the study of behavior* (Vol. 26, pp. 435-462). New York, NY: Academic Press.

Boysen, S. T. & Berntson, G. G. (1989). Numerical competence in a chimpanzee (*Pan troglodytes*). *Journal of Comparative Psychology*, *103*(1), 23-31.

Boysen, S. T. & Capaldi, E. J. (Eds.). (1993). *The development of numerical competence: Animal and human models*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Boysen, S. T. & Hallberg, K. I. (2000). Primate numerical competence: Contributions toward understanding nonhuman cognition. *Cognitive Science, 24*(3), 423-443.

Brainerd, C. J. (1979). *The origins of the number concept*. New York, NY: Praeger Publishers.

Brannon, E. M. (2002). The development of ordinal numerical knowledge in infancy. *Cognition*, *83*(3), 223-240.

Brannon, E. M., Cantlon, J. F., & Terrace, H. S. (2006). The role of reference points in ordinal numerical comparisons by rhesus macaques (*Macaca mulatta*). *Journal of Experimental Psychology: Animal Behavior Processes*, *32*(2), 120-134.

Brannon, E. M. & Roitman, J. D. (2003). Nonverbal representations of time and number in animals and human infants. In W. H. Meck (Ed.), *Functional and neural mechanisms of interval timing* (pp. 143-182). Boca Raton, FL: CRC Press.

Brannon, E. M. & Terrace, H. S. (1998). Ordering of the numerosities 1-9 by monkeys. *Science, 282*, 746-749.

Burns, B. (1992). Perceived similarity in perceptual and conceptual development: The influence of category information on perceptual organization. In B. Burns (Ed.), *Percepts, concepts and categories: The representation and processing of information* (pp. 175-231). Oxford, England: North-Holland.

Cantlon, J. F. & Brannon, E. M. (2006). Shared system for ordering small and large numbers in monkeys and humans. *Psychological Science, 17*(5), 401-406.

Cantlon, J. F. & Brannon, E. M. (2007). How much does number matter to a monkey (*Macaca mulatta*)? *Journal of Experimental Psychology: Animal Behavior Processes*, *33*(1), 32-41.

Carey, S. (2001). On the very possibility of discontinuties in conceptual development. In E. Dupoux (Ed.), *Language, brain, and cognitive development: Essays in honor of Jacques Mehler* (pp. 303-324). Cambridge, MA: The MIT Press.

Catania, A. C. (1998). The taxonomy of verbal behavior. In K. A. Lattal & M. Perone (Eds.), *Handbook of research methods in human operant behavior* (pp. 405-434). New York, NY: Springer.

Church, R. M. & Meck, W. H. (1984). The numerical attribute of stimuli. In H. L. Roitblat, T. G. Bever & H. S. Terrace (Eds.), *Animal Cognition* (pp. 445-464). Hillsdale, NJ: Lawrence Erlbaum Associates.

Clayton, M. C. & Hayes, L. J. (1999). Conceptual differences in the analysis of stimulus equivalence. *The Psychological Record*, *49*(1), 145.

Colombo, J. (Ed.). (1993). *Infant cognition: Predicting later intellectual functioning* (Vol. 5). Newbury Park, CA: Sage Publications.

Conant, L. L. (1896). *The number concept: Its origin and development*. New York, NY: Macmillan.

Cooper, R. G., Jr. (1984). Early number development: Discovering number space with addition and subtraction. In C. Sophian (Ed.), *Origins of cognitive skills* (pp. 157-192). Hillsdale, NJ: Lawrence Erlbaum Associates.

Czerny, P. & Thomas, R. K. (1975). Sameness-difference judgments in *Saimiri sciureus* based on volumetric cues. *Animal Learning & Behavior, 3*(4), 375-379.

Dantzig, T. (1939). Number, the language of science; a critical survey written for the cultured non-mathematician (3rd ed.). New York, NY: Macmillan.

Davis, H. & Memmott, J. (1982). Counting behavior in animals: A critical evaluation. *Psychological Bulletin, 92*(3), 547-571.

Davis, H. & Pérusse, R. (1988). Numerical competence in animals: Definitional issues, current evidence, and a new research agenda. *Behavioral & Brain Sciences, 11*(4), 561-615.

Davis, R. T., Leary, R. W., Stevens, D. A., & Thompson, R. F. (1967). Learning and perception of oddity problems by lemurs and seven species of monkey. *Primates, 8*(4), 311-322.

De Lillo, C. (1996). The serial organisation of behaviour by non-human primates: An evaluation of experimental paradigms. *Behavioural Brain Research, 81*(1), 1-17.

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition, 44*, 1-42.

Dehaene, S. (1997). *The number sense: How the mind creates mathematics*. New York, NY: Oxford University Press.

Dehaene, S. & Changeux, J.-P. (1993). Development of elementary numerical abilities: A neuronal model. *Journal of Cognitive Neuroscience, 5*(4), 390-407.

Dehaene, S., Dehaene-Lambertz, G., & Cohen, L. (1998). Abstract representations of numbers in the animal and human brain. *Trends in Neurosciences, 21*(8), 355-361.

Detlefsen, M. (2005). Formalism. In S. Shapiro (Ed.), *The oxford handbook of philosophy of mathematics and logic* (pp. 236-317). New York, NY: Oxford University Press.

Dooley, G. B. & Gill, T. (1977a). Acquisition and use of mathematical skills by a linguistic chimpanzee. In D. M. Rumbaugh (Ed.), *Language learning by a chimpanzee: The Lana Project* (pp. 247-260). New York, NY: Academic Press.

Dooley, G. B. & Gill, T. (1977b). Mathematical capabilities of Lana chimpanzee. In G. H. Bourne (Ed.), *Progress in ape research* (pp. 133-142). New York, NY: Academic Press.

Douglass, H. R. (1925). The development of number concept in children of pre-school and kindergarten ages. *Journal of Experimental Psychology, 8,* 443-470.

Doumas, L. A. A., Hummel, J. E., & Sandhofer, C. M. (2008). A theory of the discovery and predication of relational concepts. *Psychological Review, 115*(1), 1-43.

Draper, W. A. (1965). Cue dominance in oddity discriminations by rhesus monkeys. *Journal of Comparative and Physiological Psychology, 60*(1), 140-141.

Evans, P. M. & Smith, L. B. (1988). The development of identity as a privileged relation in classification: When very similar is just not similar enough. *Cognitive Development, 3*(3), 265-284.

Fabre-Thorpe, M. (2001). Visual categorization: Accessing abstraction in non-human primates. *Philosophical Transactions of the Royal Society of London: Series B, Biological Sciences, 358*(1435), 1215-1223.

Fagot, J., Kruschke, J. K., Dépy, D., & Vauclair, J. (1998). Associative learning in baboons (Papio papio) and humans (Homo sapiens): Species differences in learned attention to visual features. *Animal Cognition, 1*(2), 123-133.

Ferster, C. B. (1960). Intermittent reinforcement of matching to sample in the pigeon. *Journal of the Experimental Analysis of Behavior, 3,* 259-272.

Ferster, C. B. (1964). Arithmetic behavior in chimpanzees. *Scientific American, 210*(5), 98-106.

Ferster, C. B. & Hammer, C. E. (1966). Synthesizing the components of arithmetic behavior. In W. K. Honig (Ed.), *Operant behavior: Areas of research and application* (pp. 634-677). New York, NY: Appleton-Century-Crofts.

Fields, L. & Reeve, K. F. (2000). Synthesizing equivalence classes and natural categories from perceptual and relational classes. In J. C. Leslie & D. Blackman (Eds.), *Experimental and applied analysis of human behavior* (pp. 59-83). Reno, NV: Context Press.

Fujita, K. (1982). An analysis of stimulus control in two-color matching-to-sample behaviors of Japanese monkeys (Macaca fuscata fuscata). *Japanese Psychological Research, 24*(3), 124-135.

Fujita, K. (1983a). Acquisition and transfer of a higher-order conditional discrimination performance in the Japanese monkey. *Japanese Psychological Research*, *25*(1), 1-8.

Fujita, K. (1983b). Formation of the sameness–difference concept by Japanese monkeys from a small number of color stimuli. *Journal of the Experimental Analysis of Behavior*, *40*(3), 289-300.

Fujita, K. (1985). Effects of ratio reinforcement schedules on discrimination performance by Japanese monkeys. *Journal of the Experimental Analysis of Behavior*, *43*(2), 225-234.

Fuson, K. (1988) Children's counting and concepts of number. *Springer Series in Cognitive Development*. New York, NY: Springer-Verlag.

Gallistel, C. R. (1989). Animal cognition: The representation of space, time and number. *Annual Review of Psychology, 40*, 155.

Gallistel, C. R. (1993). A conceptual framework for the study of numerical estimation and arithmetic reasoning in animals. In S. T. Boysen & E. J. Capaldi (Eds.), *The development of numerical competence: Animal and human models* (pp. 211-223). Hillsdale, NJ: Lawrence Erlbaum Associates.

Gallistel, C. R. & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition*, *44*(1), 43-74.

Gallistel, C. R. & Gelman, R. (2000). Non-verbal numerical cognition: From reals to integers. *Trends in Cognitive Sciences, 4*(2), 59-65.

Galvão, O. D. F., Barros, R. D. S., Dos Santos, J. R., Brino, A. L. D., Brandão, S., Lavratti, C. M., . . . McIlvane, W. J. (2005). Extent and limits of the matching concept in *Cebus apella*: A matter of experimental control? *The Psychological Record, 55*(2), 219-232.

Garcha, H. S. & Ettlinger, G. (1979). Object sorting by chimpanzees and monkeys. *Cortex, 15*(2), 213-224.

Geary, D. C. (2000). From infancy to adulthood: The development of numerical abilities. *European Child & Adolescent Psychiatry, 9*(Suppl 2), S11-S16.

Geary, D. C. & Lin, J. (1998). Numerical cognition: Age-related differences in the speed of executing biologically primary and biologically secondary processes. *Experimental Aging Research, 24*(2), 101-137.

Gelman, R. & Gallistel, C. R. (1986). *The child's understanding of number* (2nd ed.). Cambridge, MA: Harvard University Press.

Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. *Psychological Review, 84*(3), 279-325.

Gibbon, J. & Meck, W. H. (1984). Sources of variance in an information processing theory of timing. In H. L. Roitblat, T. G. Bever & H. S. Terrace (Eds.), *Animal Cognition* (pp. 465-488). Hillsdale, NJ: Lawrence Erlbaum Associates.

Gillan, D. J., Premack, D., & Woodruff, G. (1981). Reasoning in the chimpanzee: I. Analogical reasoning. *Journal of Experimental Psychology: Animal Behavior Processes, 7*(1), 1-17.

Green, G. & Saunders, R. R. (1998). Stimulus equivalence. In K. A. Lattal & M. Perone (Eds.), *Handbook of research methods in human operant behavior* (pp. 229-262). New York, NY: Springer.

Green, G., Stromer, R., & Mackay, H. A. (1993). Relational learning in stimulus sequences. *The Psychological Record, 43*(4), 599-615.

Hamilton, A. G. (1982). *Numbers, sets, and axioms: The apparatus of mathematics*. New York, NY: Cambridge University Press.

Hashiya, K. & Kojima, S. (1997). Auditory-visual intermodal matching by a chimpanzee (*Pan troglodytes*). *Japanese Psychological Research, 39*(3), 182-190.

Haun, D. B. M., Jordan, F. M., Vallortigara, G., & Clayton, N. S. (2010). Origins of spatial, temporal and numerical cognition: Insights from comparative psychology. *Trends in Cognitive Sciences, 14*(12), 552-560.

Hauser, M. & Carey, S. (1998). Building a cognitive creature from a set of primitives: Evolutionary and developmental insights. In D. D. Cummins & C. Allen (Eds.), *Evolution of mind* (pp. 51-106). New York, NY: Oxford University Press.

Hauser, M. D. (1997). Tinkering with minds from the past. In G. R. Brock & G. Cardew (Eds.), *Characterizing human psychological adaptations* (pp. 95-131). Chichester, NY: John Wiley & Sons.

Hayes, S. C., Barnes-Holmes, D., & Roche, B. (2001). *Relational frame theory: A post-Skinnerian account of human language and cognition*. New York, NY: Kluwer Academic/Plenum Publishers.

Hayne, H., Rovee-Collier, C., & Perris, E. E. (1987). Categorization and memory retrieval by three-month-olds. *Child Development, 58*(3), 750-767.

Henry, D. E. (1976). Interrelationships among attentional preferences, cardinal-ordinal ability, and conservation of number. *Child Development, 47*(3), 750-758.

Herrnstein, R. J. (1984). Objects, categories, and discriminative stimuli. In H. L. Roitblat, T. G. Bever & H. S. Terrace (Eds.), *Animal Cognition* (pp. 233-261). Hillsdale, NJ: Lawrence Erlbaum Associates.

Herrnstein, R. J. (1990). Levels of stimulus control: A functional approach. *Cognition, 37*(1), 133-166.

Himes, G. T. (1999). *A chronometric analysis of same/different judgements based on quantity by chimpanzees (Pan troglodytes).* Unpublished master's thesis, Ohio State University, Columbus, OH.

Holyoak, K. J., Gentner, D., & Kokinov, B. N. (2001). Introduction: The place of analogy in cognition. In D. Gentner, K. J. Holyoak & B. N. Kokinov (Eds.), *The analogical mind: Perspectives from cognitive science* (pp. 1-19). Cambridge, MA: The MIT Press.

Jackson, W. J. & Pegram, G. V. (1970). Comparison of intra- vs extradimensional transfer of matching by rhesus monkeys. *Psychonomic Science*, *19*(3), 162-163.

James, W. (1890/1981). *The principles of psychology* (Vol. 1). Cambridge, MA: Harvard University Press. (Original work published 1890).

Jitsumori, M. & Delius, J. D. (2001). Object recognition and object categorization in animals. In T. Matsuzawa (Ed.), *Primate origins of human cognition and behavior* (pp. 269-293). New York, NY: Springer-Verlag Publishing.

Jitsumori, M., Siemann, M., Lehr, M., & Delius, J. D. (2002). A new approach to the formation of equivalence classes in pigeons. *Journal of the Experimental Analysis of Behavior, 78*(3), 397-408.

Jordan, K. E. & Brannon, E. M. (2006). Weber's Law influences numerical representations in rhesus macaques (*Macaca mulatta*). *Animal Cognition, 9*(3), 159-172.

Judge, P. G., Evans, T. A., & Vyas, D. K. (2005). Ordinal representation of numeric quantities by brown capuchin monkeys (*Cebus apella*). *Journal of Experimental Psychology: Animal Behavior Processes, 31*(1), 79-94.

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology, 24*(2), 175-219.

Katz, J. S., Wright, A. A., & Bodily, K. D. (2007). Issues in the comparative cognition of abstract-concept learning. *Comparative Cognition & Behavior Reviews, 2*, 79-92.

Keller, F. S. & Schoenfeld, W. N. (1950). *Principles of psychology*. New York, NY: Appleton-Century-Crofts.

Kemler, D. G. (1983). Holistic and analytic modes in perceptual and cognitive development. In T. J. Tighe & B. E. Shepp (Eds.), *Perception, cognition, and development: Interactional analyses* (pp. 77-102). Hillsdale, NJ: L. Erlbaum Associates.

King, J. E. (1973). Learning and generalization of a two-dimensional sameness-difference concept by chimpanzees and orangutans. *Journal of Comparative and Physiological Psychology, 84*(1), 140-148.

Kingma, J. & Koops, W. (1981). On the sequentiality of ordinality and cardinality. *International Journal of Behavioral Development, 4*(4), 391.

Kojima, T. (1979). Discriminative stimulus context in matching-to-sample of Japanese monkeys. *Japanese Psychological Research, 21*(4), 189-194.

Kojima, T. (1982). Discriminative stimulus context in matching-to-sample of Japanese monkeys: A further examination. *Japanese Psychological Research, 24*(3), 155-160.

Kotovsky, L. & Gentner, D. (1996). Comparison and categorization in the development of relational similarity. *Child Development, 67*(6), 2797-2822.

Kruschke, J. K. (1996). Dimensional relevance shifts in category learning. *Connection Science, 8*(2), 225-247.

Lamb, M. E., Bornstein, M. H., & Teti, D. M. (2002). *Development in infancy: An introduction* (4th ed.). Mahwah, NJ: Lawrence Erlbaum.

Lattal, K. A. & Perone, M. (1998). The experimental analysis of human operant behavior. In K. A. Lattal & M. Perone (Eds.), *Handbook of research methods in human operant behavior* (pp. 3-14). New York, NY: Springer.

Le Corre, M. & Carey, S. (2007). One, two, three, four, nothing more: An investigation of the conceptual sources of the verbal counting principles. *Cognition, 105*(2), 395-438.

Lea, S. E. G. (1984). In what sense do pigeons learn concepts? In H. L. Roitblat, T. G. Bever & H. S. Terrace (Eds.), *Animal Cognition* (pp. 263-276). Hillsdale, NJ: Lawrence Erlbaum Associates.

Lea, S. E. G. & Wills, A. J. (2008). Use of multiple dimensions in learned discriminations. *Comparative Cognition & Behavior Review, 3*, 115-133.

Levy, A. (1979). *Basic set theory*. New York, NY: Springer-Verlag Publishing.

Lock, A. & Colombo, M. (1996). Cognitive abilities in a comparative perspective. In A. Lock & C. R. Peters (Eds.), *Handbook of human symbolic evolution* (pp. 596-643). New York, NY: Oxford University Press.

Mandler, G. (1985). *Cognitive psychology: An essay in cognitive science*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Mandler, J. M. (2000). Perceptual and conceptual processes in infancy. *Journal of Cognition and Development, 1*(1), 3-36.

Mandler, J. M. (2004). Thought before language. *Trends in Cognitive Sciences, 8*(11), 508-513.

Matsuzawa, T. (1985). Use of numbers by a chimpanzee. *Nature, 315*, 57-59.

Matsuzawa, T., Asano, T., Kubota, K., & Murofushi, K. (1986). Acquisition and generalization of numerical labeling by a chimpanzee. In D. M. Taub & F. A. King (Eds.), *Current perspectives in primate social dynamics* (pp. 416-430). New York, NY: Van Nostrand Reinhold.

McMurray, B. & Aslin, R. N. (2004). Anticipatory eye movements reveal infants' auditory and visual categories. *Infancy, 6*(2), 203-229.

Meck, W. H. & Church, R. M. (1983). A mode control model of counting and timing processes. *Journal of Experimental Psychology: Animal Behavior Processes, 9*(3), 320-334.

Medin, D. L. (1989). Concepts and conceptual structure. *American Psychologist, 44*(12), 1469-1481.

Menyuk, P., Liebergott, J. W., & Schultz, M. C. (1995). *Early language development in full-term and premature infants*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Menzel, E. W., Jr. & Menzel, C. R. (2007). Do primates plan routes? Simple detour problems reconsidered. In D. A. Washburn (Ed.), *Primate perspectives on behavior and cognition* (pp. 175-206). Washington, DC: American Psychological Association.

Merritt, D. J., Rugani, R., & Brannon, E. M. (2009). Empty sets as part of the numerical continuum: Conceptual precursors to the zero concept in rhesus monkeys. *Journal of Experimental Psychology: General, 138*(2), 258-269.

Meyer, D. R. & Harlow, H. F. (1949). The development of transfer of response to patterning by monkeys. *Journal of Comparative and Physiological Psychology, 42*(6), 454-462.

Michell, J. (1997). Quantitative science and the definition of measurement in psychology. *British Journal of Psychology, 88*(3), 355-383.

Murofushi, K. (1997). Numerical matching behavior by a chimpanzee (*Pan troglodytes*): Subitizing and analogue magnitude estimation. *Japanese Psychological Research, 39*(3), 140-153.

Nakagawa, E. (2003). Shift learning in same-different conditional discriminations in rats. *The Psychological Record, 53*(3), 487-506.

Neisser, U. (1987). From direct perception to conceptual structure. In U. Neisser (Ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization* (pp. 11-24). New York, NY: Cambridge University Press.

Nissen, H. W., Blum, J. S., & Blum, R. A. (1949). Conditional matching behavior in chimpanzee; implications for the comparative study of intelligence. *Journal of Comparative and Physiological Psychology, 42*(5), 339-356.

Noble, L. M. & Thomas, R. K. (1985). Oddity and dimension-abstracted oddity (DAO) in humans. *American Journal of Psychology, 98*(4), 549-557.

Oden, D. L., Thompson, R. K., & Premack, D. (1988). Spontaneous transfer of matching by infant chimpanzees (*Pan troglodytes*). *Journal of Experimental Psychology: Animal Behavior Processes, 14*(2), 140-145.

Olthof, A., Iden, C. M., & Roberts, W. A. (1997). Judgments of ordinality and summation of number symbols by squirrel monkeys (*Saimiri sciureus*). *Journal of Experimental Psychology: Animal Behavior Processes, 23*(3), 325-339.

Olthof, A. & Roberts, W. A. (2000). Summation of symbols by pigeons (*Columba livia*): The importance of number and mass of reward items. *Journal of Comparative Psychology, 114*(2), 158-166.

Palmer, D. C. (2002). Psychological essentialism: A review of E. Margolis and S. Laurence (EDS.), Concepts: Core readings. *Journal of the Experimental Analysis of Behavior, 78*(3), 597-607.

Parker, S. T. (1999). The life history and development of great apes in comparative perspective. In S. T. Parker, R. W. Mitchell & H. L. Miles (Eds.), *The mentalities of gorillas and orangutans: Comparative perspectives* (pp. 43-69). New York, NY: Cambridge University Press.

Parker, S. T., Mitchell, R. W., & Miles, H. L. (Eds.). (1999). *The mentalities of gorillas and orangutans: Comparative perspectives*. New York, NY: Cambridge University Press.

Pearce, J. M. (1988). Stimulus generalization and the acquisition of categories by pigeons. In L. Weiskrantz (Ed.), *Thought without language* (pp. 132-155). New York, NY: Clarendon Press/Oxford University Press.

Piaget, J. (1941/1965). *The child's conception of number*. New York, NY: W.W. Norton & Company.

Piek, J. P. (2006). *Infant motor development*. Champaign, IL: Human Kinetics.

Premack, D. (1988). Minds with and without language. In L. Weiskrantz (Ed.), *Thought without language* (pp. 46-65). New York, NY: Oxford University Press.

Roberts, A. C., Robbins, T. W., & Everitt, B. J. (1988). The effects of intradimensional and extradimensional shifts on visual discrimination learning in humans and non-human primates. *The Quarterly Journal of Experimental Psychology B: Comparative and Physiological Psychology, 40*(4), 321-341.

Robinson, J. S. (1955). The sameness-difference discrimination problem in chimpanzee. *Journal of Comparative and Physiological Psychology, 48*, 195-197.

Rovee-Collier, C. (1996). Measuring infant memory: A critical commentary. *Developmental Review, 16*(3), 301-310.

Saunders, K. J. & Williams, D. C. (1998). Stimulus-control procedures. In K. A. Lattal & M. Perone (Eds.), *Handbook of research methods in human operant behavior* (pp. 193-228). New York, NY: Springer.

Sidman, M. (1990). Equivalence relations: Where do they come from? In H. Lejeune & D. E. Blackman (Eds.), *Behaviour analysis in theory and practice: Contributions and controversies* (pp. 93-114). Hillsdale, NJ: Lawrence Erlbaum Associates.

Sidman, M. (1997). Equivalence relations. *Journal of the Experimental Analysis of Behavior, 68*(2), 258-266.

Sidman, M., Wynne, C. K., Maguire, R. W., & Barnes, T. (1989). Functional classes and equivalence relations. *Journal of the Experimental Analysis of Behavior, 52*(3), 261-274.

Siegel, L. S. (1974). Development of number concepts: Ordering and correspondence operations and the role of length cues. *Developmental Psychology, 10*(6), 907-912.

Simon, T. J. (1997). Reconceptualizing the origins of number knowledge: A "non-numerical" account. *Cognitive Development, 12*(3), 349-372.

Smith, E. E. (1981) Categories and concepts. In D. N. Osherson (Series Ed.), *An invitation to cognitive science: Vol. 3. Thinking* (2nd ed., pp. 3-33). Cambridge, MA: Harvard University Press.

Smith, L. B. (1984). Young children's understanding of attributes and dimensions: A comparison of conceptual and linguistic measures. *Child Development, 55*(2), 363-380.

Smith, L. B. (1989). From global similarities to kinds of similarities: The construction of dimensions in development. In S. Vosniadou & O. Andrew (Eds.), *Similarity and analogical reasoning* (pp. 215-252). New York, NY: Cambridge University Press.

Smith, L. B. (1993). The concept of same. *Advances in Child Development and Behavior, 24*, 215-252.

Smith, L. B. (2005). Emerging ideas about categories. In L. Gershkoff-Stowe & D. H. Rakison (Eds.), *Building object categories in developmental time* (pp. 159-173). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Smith, L. B. & Heise, D. (1992). Perceptual similarity and conceptual structure. In B. Burns (Ed.), *Percepts, concepts and categories: The representation and processing of information* (pp. 233-272). Oxford, England: North-Holland.

Spinozzi, G. (1996). Categorization in monkeys and chimpanzees. *Behavioural Brain Research, 74*(1), 17-24.

Steirn, J. N. & Thomas, R. K. (1990). Comparative assessments of intelligence: Performances of *Homo sapiens sapiens* on hierarchies of oddity and sameness-difference tasks. *Journal of Comparative Psychology, 104*(4), 326-333.

Stevens, S. S. (1951). Mathematics, measurement, and psychophysics. In S. S. Stevens (Ed.), *Handbook of experimental psychology* (pp. 1-49). New York, NY: John Wiley & Sons.

Strauss, M. S. & Curtis, L. E. (1984). Development of numerical concepts in infancy. In C. Sophian (Ed.), *Origins of cognitive skills* (pp. 131-155). Hillsdale, NJ: Lawrence Erlbaum Associates.

Strong, P. N., Jr., Drash, P., & Hedges, M. (1968). Solution of dimension abstracted oddity as function of species, experience, and intelligence. *Psychonomic Science, 11*(9), 337-338.

Terrace, H. S. (1984). Animal cognition. In H. L. Roitblat, T. G. Bever & H. S. Terrace (Eds.), *Animal Cognition* (pp. 7-28). Hillsdale, NJ: Lawrence Erlbaum Associates.

Thagard, P. (1996). *Mind: Introduction to cognitive science*. Cambridge, MA: The MIT Press.

Thomas, R. K. (1988). To honor of Davis & Pérusse and repeal their glossary of processes of numerical competence. *Behavioral & Brain Sciences, 11*(4), 600.

Thomas, R. K. & Frost, T. (1983). Oddity and dimension-abstracted oddity (DAO) in squirrel monkeys. *American Journal of Psychology, 96*(1), 51-64.

Thomas, R. K. & Peay, L. (1976). Length judgments by squirrel monkeys: Evidence for conservation? *Developmental Psychology, 12*(4), 349-352.

Thompson, R. K. R. (1995). Natural and relational concepts in animals. In H. L. Roitblat & J. A. Meyer (Eds.), *Comparative approaches to cognitive science* (pp. 175-224). Cambridge, MA: The MIT Press.

Thompson, R. K. R. & Oden, D. L. (1995). A profound disparity revisited: Perception and judgment of abstract identity relations by chimpanzees, human infants, and monkeys. *Behavioural Processes, 35*(1-3), 149-161.

Thompson, R. K. R. & Oden, D. L. (2000). Categorical perception and conceptual judgments by nonhuman primates: The paleological monkey and the analogical ape. *Cognitive Science, 24*(3), 363-396.

Tomasello, M. (1999). *The cultural origins of human cognition*. Cambridge, MA: Harvard University Press.

Tomasello, M. & Call, J. (1997). *Primate cognition*. London, UK: Oxford University Press.

Tomonaga, M. (1993). Tests for control by exclusion and negative stimulus relations of arbitrary matching to sample in a "symmetry-emergent" chimpanzee. *Journal of the Experimental Analysis of Behavior, 59*(1), 215-229.

Tomonaga, M. (1995). Transfer of odd-item search performance in a chimpanzee (*Pan troglodytes*). *Perceptual and Motor Skills, 80*(1), 35-42.

Truppa, V., Garofoli, D., Castorina, G., Mortari, E. P., Natale, F., & Visalberghi, E. (2010). Identity concept learning in matching-to-sample tasks by tufted capuchin monkeys (*Cebus apella*). *Animal Cognition, 13*(6), 835-848.

Urcuioli, P. J. & Zentall, T. R. (1993). A test of comparison-stimulus substitutability follow one-to-many-matching by pigeons. *The Psychological Record, 43*(4), 745-759.

Vasconcelos, M. (2008). Transitive inference in non-human animals: An empirical and theoretical analysis. *Behavioural Processes, 78*(3), 313-334.

Vonk, J. (2003). Gorilla (*Gorilla gorilla gorilla*) and orangutan (*Pongo abelii*) understanding of first- and second-order relations. *Animal Cognition, 6*(2), 77-86.

Vonk, J. & MacDonald, S. E. (2002). Natural concepts in a juvenile gorilla (*Gorilla gorilla gorilla*) at three levels of abstraction. *Journal of the Experimental Analysis of Behavior, 78*(3), 315-329.

Wasserman, E. A. & DeVolder, C. L. (1993). Similarity- and nonsimilarity-based conceptualization in children and pigeons. *The Psychological Record, 43*(4), 779-793.

Wasserman, E. A., Kiedinger, R. E., & Bhatt, R. S. (1988). Conceptual behavior in pigeons: Categories, subcategories, and pseudocategories. *Journal of Experimental Psychology: Animal Behavior Processes, 14*(3), 235-246.

Wasserman, E. A. & Young, M. E. (2010). Same–different discrimination: The keel and backbone of thought and reasoning. *Journal of Experimental Psychology: Animal Behavior Processes, 36*(1), 3-22.

Weinstein, B. (1945). The evolution of intelligent behavior in rhesus monkeys. *Genetic Psychology Monographs, 31*, 3-48.

Weisberg, P. & Rovee-Collier, C. (1998). Behavioral processes of infants and young children. In K. A. Lattal & M. Perone (Eds.), *Handbook of research methods in human operant behavior* (pp. 325-370). New York, NY: Springer.

Whitehead, A. N. & Russell, B. (1927). *Principia mathematica* (2nd ed.). New York, NY: Cambridge University Press.

Woodruff, G. & Premack, D. (1981). Primative mathematical concepts in the chimpanzee: Proportionality and numerosity. *Nature, 293*(5833), 568-570.

Wright, A. A. (1992). Testing the cognitive capacities of animals. In I. Gormezano & E. A. Wasserman (Eds.), *Learning and memory: The behavioral and biological substrates* (pp. 45-60). Hillsdale, NJ: Lawrence Erlbaum Associates.

Wright, A. A. & Katz, J. S. (2006). Mechanisms of same/different concept learning in primates and avians. *Behavioural Processes, 72*(2006), 234-254.

Wright, A. A. & Katz, J. S. (2009). A case for restricted-domain relational learning. *Psychonomic Bulletin & Review, 16*(5), 907-913.

Wright, A. A. & Lickteig, M. T. (2010). What is learned when concept learning fails?—A theory of restricted-domain relational learning. *Learning and Motivation, 41*(4), 273-286.

Wynn, K. (1992a). Evidence against empiricist accounts of the origins of numerical knowledge. *Mind & Language, 7*(4), 315-332.

Wynn, K. (1992b). Issues concerning a nativist theory of numerical knowledge. *Mind & Language, 7*(4), 367-381.

Wynn, K. (1998a). An evolved capacity for number. In D. D. Cummins & C. Allen (Eds.), *Evolution of mind* (pp. 107-126). New York, NY: Oxford University Press.

Wynn, K. (1998b). Psychological foundations of number: Numerical comptence in human infants. *Trends in Cognitive Sciences, 2*(8), 296-303.

Yagi, B., Shinoda, A., Shinohara, S., & Hirata, A. (1975). Analysis of color-form cue problem in the Japanese monkey (*Macaca fuscata yakui*): I. Oddity problem [Abstract]. *Annual of Animal Psychology, 24*(2), 87-98.

Young, M. L. & Harlow, H. F. (1943). Generalization by rhesus monkeys of a problem involving the Weigl principle using the oddity method. *Journal of Comparative Psychology, 36*(3), 201-216.

Zentall, T. R. (1996). An analysis of stimulus class formation in animals. In T. R. Zentall & P. M. Smeets (Eds.), *Stimulus class formation in humans and animals* (pp. 15-34). New York, NY: Elsevier Science Publishers.

Zentall, T. R., Clement, T. S., & Weaver, J. E. (2003). Symmetry training in pigeons can produce functional equivalences. *Psychonomic Bulletin & Review, 10*(2), 387-391.

Zentall, T. R., Galizio, M., & Critchfield, T. S. (2002). Categorization, concept learning and behavior analysis: An introduction. *Journal of the Experimental Analysis of Behavior, 78*(3), 237-248.

Zentall, T. R., Hogan, D. E., & Edwards, C. E. (1984). Cognitive factors in conditional learning by pigeons. In H. L. Roitblat, T. G. Bever & H. S. Terrace (Eds.), *Animal Cognition* (pp. 389-405). Hillsdale, NJ: Lawrence Erlbaum Associates.

Zentall, T. R. & Smeets, P. M. (1996). Preface. In T. R. Zentall & P. M. Smeets (Eds.), *Stimulus class formation in humans and animals* (pp. v-viii). New York, NY: Elsevier Science Publishers.

Zentall, T. R., Wasserman, E. A., Lazareva, O. F., Thompson, R. K. R., & Rattermann, M. J. (2008). Concept learning in animals. *Comparative Cognition & Behavior Reviews, 3,* 13-45.

Zimiles, H. (1963). A note on Piaget's concept of conservation. *Child Development, 34*(3), 691-695.